A *Nondeterministically Enumerated Categorial Grammar* Analysis Of Croatian And English Passive Constructions

Undergraduate Research Thesis

Presented in Partial Fulfillment of the Requirements for graduation "with Research Distinction in Linguistics" in the undergraduate colleges of The Ohio State University

by
Michael James Sullivan

The Ohio State University
May 2019

Project Advisors: Professor Robert Levine and Professor Andrea Sims, Department of Linguistics

Table of Contents

(1) Introduction

Passive or passive like constructions, such as those in the below Croatian data, make regular appearances in the cross-linguistic topography (Dixon & Aikhenvald 2010). These morphosyntactic phenomena reduce valence and promote a canonical object to the canonical subject position or inflection, or both. Their inherent complexity presents formal challenges to any current syntactic framework, and yet, as salient and widespread linguistic data, they must be accurately accounted for. This paper will introduce relevant data from three separate classes of modern Croatian passive and passive-like constructions. The creation of a formal explanation for these phenomena presents several interesting challenges to any theoretical framework, due to several morphological, syntactic, and semantic idiosyncrasies, whose variation between constructions seems to defy any sort of apparent pattern.

The objective of this thesis is to explain the obstacles to an analysis in both Transformational (TG) and Categorial Grammars (CG), and finally, analyze the data in the *Nondeterministically Enumerated Categorial Grammar* (NECG) framework, while simultaneously deriving (nearly) synonymous English examples and providing the reader with a step-by-step explanation of the derivational mechanisms of NECG.

(1.1) The Croatian Data

The following section will provide examples of, and describe, four different Croatian morphosyntactic constructions; active (1), passive (2), impersonal (3), and deverbal adjectival (4). In the following section, four corresponding near-synonymous English constructions will be given (1b-4b), one for each of the Croatian examples.

Each example is meant to be as close to a minimal pair as possible with the active constructions given in (1) and (1b)-- the main verb, ZATVORITI (*to close*), is the same across all four examples, the semantic arguments are kept the same where possible (MARKO will always be the agent-like argument, when an agent is possible, and KNJIGA will always be the patient-like argument), and the tense/aspect is kept as invariant as is feasible. Beginning with the active construction (here, as in the rest of the paper, all morphosyntactic values are given in standard Leipzig Glossing notation-- Comrie et al. 2010);

1. (*Active*)

Marko	je	zatvori-o	knjig-u
Marko-NOM	AUX.3.SG	shut-PTCP.M.SG	book(F)-ACC.SG

"Marko shut a book"

In this construction, note that the verb ZATVORITI is in its participial form (*zatvorio*), and agrees in number (singular) and gender (masculine) with the nominative subject, *Marko*. As one would expect, the object, *knjigu* is marked with accusative inflection. Note that Croatian word order is relatively free, meaning that *Marko*, *zatvorio*, and *knjigu* may grammatically appear in any order (Snježana 1997). The present third person singular auxiliary-- agreeing with the subject *Marko* in person and number-- *je* (a finite form of the verb BITI), however, is a clitic, and as such is required to appear following the first phrase; in this case the NP *Marko*. This construction yields the following semantic interpretation; $\exists(\mathbf{book})(\lambda x.close(x)(marko)))$

There are several notable morphosyntactic variations between an active construction and its corresponding passive counterpart;

2. (*Passive*)

Zatvori-la	se+[je]	knjig-a	(od	Mark-a)
shut-PTCP.F.SG	SE+[AUX.3.SG]	book(F)-NOM.SG	(by	Marko-GEN)

"A book was shut (by Marko)"

Most saliently, note that the "se+[je]" in the Croatian portion of the gloss indicates that the only overtly realized clitic is SE (a third person reflexive/impersonal pronoun), and that the auxiliary *je* does not actually appear phonologically, meaning that the passive sentence in (2) reads; "*zatvorila se kniga* (*od Marka*)". *je* is left in brackets to indicate that this auxiliary clitic is often analysed in this context as underlyingly present, but lacking an overt phonological realization (Zec 1985). In this analysis, the absence of *je* is analyzed as a purely syntactophonological phenomenon-- see section (2.2.2) for a more in-depth explanation.

Additionally, note that the accusative object of the active sentence in (1), *knjigu* (from the noun KNJIGA-- "*book*"), appears here as a *nominative* subject, and the inflectional agreement morphology of the participial *zatvorila* matches in gender and number accordingly (feminine singular). As in English, the agent-like role that was assigned to the subject in the active form may be optionally reintroduced via a prepositional phrase. In Croatian, this preposition is OD (corresponding to the English BY), which selects for a genitive NP as its argument (e.g. "*od Marka*"). With an *od*-phrase, the passive interpretation is the same as in its active counterpart (i.e. $\exists(\mathbf{book})(\lambda x.\mathit{close}(x)(\mathit{marko})))$, and omitting the *od*-phrase results in the following similar reading; $\exists(\mathbf{book})(\lambda y.\exists(\mathbf{x})(\lambda z.\mathit{close}(y)(z)))$.

This construction has several notable differences when compared to the passive-like impersonal;

3. (*Impersonal*)

Zatvori-lo	se+[je]	knjig-u	*(od	Mark-a)
shut-PTCP.N.SG	se+[AUX.3.SG]	book(F)-ACC.SG	*(by	Marko-GEN)

"A book was shut *(by Marko)."

As in the passive, note the presence of the ostensibly pronominal clitic SE, and subsequent overt absence of the auxiliary clitic *je*. In this construction, contrasting with the other three examples in (1), (2), and (4), *there is no nominative subject*. The participial ZATVORITI instead inflects for *neuter* singular agreement morphology (*zatvorilo*), despite the lack of any neuter NP within the construction. As in the active form (1), the patient-like/canonical object argument, *knjigu*, takes accusative inflection. Unlike the passive (2), the agent-like argument *cannot* be reintroduced via a prepositional *od*-phrase in the impersonal (Spalatin 1973), and the corresponding interpretation implies that *there does not exist an agentive argument*; $\exists x \neg \exists y[\mathit{book}(x) \land \mathit{close}(x)(y)]$.

The final passive-like construction, the deverbal adjectival, more closely resembles the passive (2);

4. (*Deverbal adjectival*)

Knjig-a	je	bi-la	zatvor-en-a	(od	Mark-a)
book(F)-NOM.SG	AUX.3SG	AUX-PTCP.F.SG	shut-ADJ-F.SG	(by	Marko-GEN)

"A book was shut (by Marko)."

Here, as in the passive, the nominative NP argument (*knjiga*) is assigned the patient-like semantic role, and the nonfinite verbs agree in both gender and number (feminine singular). Again resembling the passive construction, there is optional reintroduction of the agent-like argument using a prepositional *od*-phrase. However, those two shared characteristics comprise the extent of the similarity between the deverbal adjectival construction and the passive form given in (2). The auxiliary clitic *je*, seemingly replaced by *se* in the other two valence reducing forms, the passive and impersonal, is in fact present here, and, correspondingly, the pronominal clitic *se* is notably absent.

Additionally, a verb form that is not used in any of the first three examples, *bila*, appears in (4). This is the feminine singular pronominal form of the auxiliary BITI, whose finite (third person singular) form, the clitic *je*, which (at least, overtly) only surfaces here and in the active example given in (1)-- the two constructions formed without the clitic SE. BITI is roughly synonymous with the English BE, and, as in its English counterpart, has both auxiliary (e.g. "*John is sleeping*", where BE takes a present participial argument) and copular (e.g. "*John is happy*", where BE takes an adjective as its argument) forms. Contrasting again with the examples in (1-3), ZATVORITI does not take participial inflection, but rather has *deverbal adjectival* morphology (*zatvorena*). This is similar to a class of English passive-like deverbal constructions (e.g. *closed* in "*The door was closed when Mary arrived*") which, while phonologically identical to their passive/past participial forms, crucially differ in the adjectival form's *stative* interpretation. Note the sharp contrast in grammaticality between "*The door is locked today*" and *"*The door is locked <u>by James</u> today*".

The largest difficulty in an analysis of these patterns, regardless of the framework in question, lies in accounting for the interaction between the pronominal SE and the larger constructructions in which it appears, at both the syntax-semantics and syntactophonological interfaces. In a traditional Transformational account along the lines of Baker et al. (1989), it is not unreasonable to hypothesize that BE is the Croatian analogue to the English *-en* passive morpheme (e.g. *given*) proposed in their work, which "absorbs" both the theta-role of the subject argument and the Case ("*uppercase C*"/*abstract* Case, as opposed to the "*lowercase c*" inflectional case) of the accusative Case-marking position, forcing the accusative argument to move to subject position (and consequently receive nominative Case) while retaining its patient or patient-like role.

This analysis, however, would not accurately account for the difference in both interpretation and nonfinite verbal inflection between the passive and impersonal forms. The passive morpheme is analysed as literally an argument of the verb which it modifies, subsequently absorbing the canonical subject/nominative theta-role and preventing it from being assigned to the subject NP. As previously mentioned, the impersonal form carries the interpretation that there does not exist an agent-like argument. For example, in the impersonal example given in (3); "*zatvorilo se knjigu*"/*a book was closed*, an inanimate force (e.g. the wind or gravity) may have been the cause of the book's closing, but there is the interpretation that there was no volitional, animate, or conscious entity acting as the agent in the event.

A Transformational analysis would then have to account for the fact that the same morpheme, the clitic SE, appears in both the passive and impersonal constructions, which yield different respective semantic interpretations. Additionally, the morpheme SE "swallowing" the agentive theta-role and accusative Case does not explain the obligatory neuter singular agreement inflection of the participial ZATVORITI in the impersonal, or the presence of the accusative argument, *knjigu*. Yet another question is raised regarding the optional reintroduction of the agent-like argument via an *od*-phrase in the passive, which is not a possibility in the impersonal. At best, an analysis along these lines would have to posit three homophonous forms of SE that all coincidentally have the same combinatorial syntactic properties.

The first of which being the *passive* SE, which would function identically to the English -*en* passive morpheme under the Baker et al. (1989) analysis in that it absorbs the accusative Case and agent-like theta-role, and allows for reintroduction of the absorbed role via a prepositional phrase. The second form, *impersonal* SE, imparts neuter singular agreement inflection on the nonfinite verb and absorbs the nominative Case, while additionally conferring the impersonal non-agentive interpretation. Unlike its passive counterpart, this form would not allow an *od*-phrase to obliquely reintroduce the agent-like argument. Finally, the *pronominal* SE, which appears in many constructions of the Bosnian-Croatian-Serbian (BCS) continuum as a third person reflexive and reciprocal pronoun.

Additionally, this analysis still must explain the fact that both English (2b) and Croatian (2) passive constructions optionally permit the canonical subject argument to be reintroduced through a prepositional phrase. The inescapable question remains; if the passive morpheme is an argument of the verb it modifies, and consequently assigned the agent-like theta role, by what mechanism is that same role both "unassigned" from the passive morpheme and then reassigned to the internal argument of a PP? Moreover, why is this theta-reassignment operation seemingly utilized only in those relatively few passive-like constructions that permit the oblique reintroduction of the canonical subject theta-role? At this point, the growing complexity of this analysis prompts one to ponder whether a Categorial Grammar (CG) framework-- where the mechanism that interfaces with the semantic component makes no reference to thematic roles of any sort-- could provide a more concise theoretical explanation of this Croatian passive-like data.

In comparison to the Transformationalist tradition, the Categorial/Type-Logical family of grammatical frameworks may be less familiar to a given reader. As such, the following six paragraphs provide a brief explanation of their core mechanisms, and one possessing a reasonable amount of background knowledge of this class of grammatical framework may proceed directly to the final paragraph of page 7.

Categorial Grammars (CGs) can trace their roots to Lambek (1958), and have several salient features distinguishing them from tree-based phrase structure grammars e.g. Principles and Parameters Theory and its descendents. First and foremost, CGs do not utilize branching structures, instead relying on "flat" strings which are concatenated together according to their associated syntactic types.

Additionally, these terms, rather than being *derived* in the traditional phrase structure sense, are instead *proven* in fragment Gentzen-style proof calculi, and the encoding of syntactic combinatorics is analogous to logical implication. At their core, the syntactic types of Categorial Grammars are fairly simple; there exists a finite set of *atomic types*, which can vary by specific individual framework, but for the present demonstrational purposes will be the following; S, NP, and N-- corresponding to the well-known phrasal categories *sentence*, *noun phrase*, and *noun*, respectively, and may be subtyped according to morphosyntactic values (e.g. a nominative noun phrase is of type $\mathrm{NP}_{\mathrm{NOM}}$, accusative of type $\mathrm{NP}_{\mathrm{ACC}}$, etc.) and other subcategorization metrics. From these *atomic types*, a theoretically infinite number of *complex types* may be enumerated from the following rule; S, NP and N are types, and if X is a type and Y is a type, X / Y and $Y \backslash X$ are also types.

From this set of types, syntactic categories may be assigned to lexical entries, and a given term's combinatorics are then transparently encoded as its syntactic type. For example, a finite English intransitive verb is of the type $NP_{NOM} \backslash S$ -- when concatenated with a string of type $\mathrm{NP}_{\mathrm{NOM}}$ on the left, the backward slash is *eliminated*, yielding a string of type S, which corresponds to the type of grammatical sentences. A finite transitive verb is then of the type $(NP \backslash S) / NP$, which states that the concatenation of an NP-type to the right will yield a string with the type $NP \backslash S$, which then will combine with a term of type NP on the left to yield a grammatical sentence. A ditransitive verb is then of type $(NP \backslash S) / NP / NP$, which combines with an NP-type on the right to obtain a string of the type $(NP \backslash S) / NP$, which is then syntactically identical to a transitive verb, and follows the same steps to reduce to a term of type S. Any given term whose combinatorics will eventually reduce down to the *atomic type* S will be referred to in this paper as an *S-reducing type*.

In addition to *atomic types*, a given term may select for a *complex type*. For example, an English adverb such as "*easily*" is of type $(NP \backslash S) / (NP \backslash S)$ -- it concatenates to the right of an intransitive verb, to yield a string of the same type. An example proof (i.e. derivation) of the

English sentence "*John easily won the game*" is given in figure 1 below, in order to give a more concrete example of these relatively abstract definitions:

Figure 1

the; NP / N game; N
---------------------------------- / Elim.
won; $(NP \backslash S) / NP$ the•game; NP
-- / Elim.
easily; $(NP \backslash S) / (NP \backslash S)$ won•the•game; $NP \backslash S$
-- / Elim.
easily•won•the•game; $NP \backslash S$ john; NP
-- \ Elim.
john•easily•won•the•game; S

In the above proof, note that the determiner *the* combines with the noun *game* on its right to yield an NP, while the proper noun *John* is already defined as type NP in the lexicon, and as such cannot be selected by a determiner. Additionally, it is important to mention that in each step of the proof, the left/right order of any two given terms does not necessarily correspond to the directions in which they will be concatenated. Instead, the *functor* term, i.e. the term *X* whose syntactic type selects for an *argument* term *Y*, will always appear on the left-hand side. For example, the *functor* easily•won•the•game; $NP \backslash S$ selects for its subject *argument*, john; NP, which is concatenated on the left edge, yet in the proof given in Figure 1, easily•won•the•game; $NP \backslash S$ appears on the left hand side, while john; NP is to its right.

In this instance, the syntactic terms' corresponding semantic interpretations were omitted purely for simplicity, but note that, owing to the *Curry-Howard Isomorphism* (Sørensen and Urzyczyn 2006), any given syntactic type-- a term of a directional combinatory logic-- can be assigned a corresponding term in a typed λ-calculus whose abstraction-driven combinatorics are *isomorphic* to the order of application of the directional implication, and will reduce over the course of a derivational proof to yield a logical semantic interpretation.

Before returning to the Croatian data at hand, it is important to mention that while the previous six paragraphs only describe two *directional* combinatorial rules (*left* and *right elimination*), there do exist CG analyses of languages with relatively free word orders that utilize *nondirectional* rules to account for the essentially unpredictable variation in the linear order of a given set of arguments within a construction. For further reading, see this CG analysis of Turkish in Cem (2000)

An analysis within a CG framework that could account for the morphosyntactic alternations between the active (1), passive (2), impersonal (3), and deverbal adjectival (4) constructions introduced at the beginning of this section might involve positing two terms that are homophonous, and have an identical syntactic type, to the pronominal reflexive clitic SE. One form then selects for terms with S_{IMP}-reducing syntactic types, and the other, S_{PASS}-reducing (*impersonal* and *passive*, respectively). Certain CGs, e.g. *Hybrid Type-Logical Categorial Grammar* (HTLCG), allow for *lexical rules*; in simple terms, a *lexical rule* states "given a term of type *X*, the function *F* may be applied to *X* to yield a term of type *Y* ." English passives provide a concrete example; for any given S-reducing type of the form $(NP \setminus S) / \ldots NP_n$, where $n \geq 1$ (i.e. all verbs that aren't intransitive), a passive form can be obtained by removing an NP from the term's combinatorics, applying its past participial phonological form, and rearranging the λ-abstraction operators binding the term's semantic arguments.

For example, the transitive form of the verb SEE corresponds to the following term;

5a. see; $\lambda x \lambda y.\mathit{see}(x)(y)$; $(NP \setminus S) / NP$

from which the *passivization lexical rule* can then derive a term that roughly corresponds to this;

5b. seen; $\lambda x.\mathit{see}(x)(y)$; $NP \setminus S_{PASS}$

An additional *lexical rule* also allows for the reintroduction of the outer argument (*agentive*, in Transformational vocabulary) via a *by*-phrase, which will yield the term in (5c);

5c. seen; $\lambda y \lambda x.\mathit{see}(x)(y)$; $(NP \setminus S) / PP_{BY}$

It is then relatively simple to conceive of analogous *lexical rules* that could be applied to account for the Croatian passive-like constructions given in examples (2-4).
By removing the NP_{ACC} argument, and manipulating the (presumably nondirectional) corresponding semantic combinatorics such that the NP_{NOM} argument then corresponds to the semantic variable that was previously assigned to the NP_{ACC}, the alterations in semantic interpretation and syntactic valence corresponding to the passive (2) and deverbal adjectival (4) constructions can be achieved. In the same fashion as that of the outline of the HTLCG *lexical rule* for English passivization given in (5a-c), it is then trivial to construct an optional rule which would allow both of these constructions to reintroduce the semantic argument that canonically (i.e. in the active voice) corresponds to the NP_{NOM} argument with a genitive *od*-phrase (i.e. type PP_{OD} / NP_{GEN}).

In the same vein, the impersonal form could be obtained via *lexical rule* as well, with the key difference lying in the addition of a negated existential quantifier binding the outer (i.e. canonical subject/agent-like) semantic argument. Additionally, this rule would only license participial forms with neuter singular agreement morphology, and the optional rule allowing the reintroduction of the agent-like semantic argument via a prepositional *od*-phrase would have to be sufficiently subtype-specified to prevent its application to S_{IMP}-reducing types, yet still license its application to valid S_{PASS}-reducing and deverbal adjectival types.

While this analysis-- if fully developed, of course-- could very well accurately account for the morphosyntactic variations between the constructions described in examples (1-4), it does leave much to be desired vis-à-vis *explanatory* theoretical capacity. There is no immediately obvious way to explain the fact that both the deverbal adjectival and passive constructions have essentially the same exact combinatorics regarding argument selection; both forms remove the canonical accusative NP argument, assign its semantic role to the nominative subject NP, and optionally allow for the reintroduction of the agent-like argument with an *od*-phrase. Despite these striking coincidences, these two verb forms have essentially no common morphosyntactic attributes; while the deverbal adjectival is morphologically distinguished via the *-en-* morpheme, ZATVORITI bears the same inflectional morphology in both the passive and active forms.

Additionally, the clitic SE is absent in the deverbal adjectival form, and the auxiliary clitic *je*-- which does not surface in either the passive or impersonal-- is instead present. The participial form of the copular/auxiliary verb BITI, *bila* (synonymous with the English copular past participle *been*), is present in the construction, which, as with *je*, does not appear in the passive or impersonal.

Essentially, this line of analysis necessitates the definition of three separate *lexical rules*, each one licensing a different passive-like construction from examples (2-4). The two rules from which the deverbal adjectival and passive forms are respectively obtained will each result in the exact same alterations to a given verb's canonical morphosyntactic-- and corresponding semantic-- combinatorics regarding argument selection, but then license two terms with essentially disjoint phonological and auxiliary-selectional (in a passive construction e.g. that given in example 2-- "*zatvorila se knjiga*", *zatvorila* is a nonfinite participle, yet no overt auxiliary appears in the construction) properties.

The third rule-- that which accounts for the alterations corresponding to the impersonal-- unlike its passive *and* deverbal adjectival counterparts, does not remove the accusative NP, but rather the *nominative* subject NP, and-- again, unlike the other two passive-like constructions-- does not permit the reintroduction of the agent-like semantic role via an *od*-phrase. Additionally-- contrasting with all of the other constructions given in examples (1-3)-- the participial verb obligatorily displays *neuter singular* subject agreement inflection, in spite of the conspicuous lack not only of a neuter singular NP argument, but in fact of any subject NP argument at all. Further, the semantic interpretation yielded by this hypothetical rule must confer the requisite non-agentive interpretation, which is unique to the impersonal with respect to the other constructions given in the beginning of this section.

Despite these salient contrasts, both the impersonal and the passive forms share one common characteristic; they both require the clitic SE, which is employed as a third person reflexive pronoun in certain other classes of BCS constructions. This apparent pronoun then seems to block the overt surfacing of the auxiliary *je*, which is notably *not* generally analysed as non-existent, due to the fact that the both passive and impersonal forms of ZATVORITI in examples (2) and (4) are nonfinite participles, yet no auxiliary, finite or otherwise, is present in the construction.

All of this then naturally begs the question for which this *lexical rule*-based analysis cannot produce a satisfying answer; what grammatical "work", syntactic, morphological, semantic, or some combination of the three, is the SE clitic performing? Why does the Croatian language require that it be present in either the passive or impersonal constructions? One might be tempted to posit that it's effect is morphosyntactic, somehow causing the verb's agreement inflection to permit the lack of a nominative NP and neuter singular verbal inflection inherent to the impersonal. While reasonable, this hypothesis ignores the fact that SE is present in the passive construction as well, without interfering with the feminine singular agreement inflection on the participle *zatvorila* in example (2), which agrees as expected with the nominative subject NP *knjiga*.

(1.2) The English Data

Here, a near-synonymous English construction will be given for each of the four Croatian examples (1-4) introduced in the previous section:

1b. (*Active*). John closed a book.
2b. A book was closed (by John).
3b. A book closed *(by John) (when Mary opened the window).
4b. A book was closed *(by John) (when Mary entered the room).

While English does not have the prolific noun-case inflection-- or permit the fluid word order-- as observed in Croatian, clear parallels can be drawn between the two sets of data. The English passive (2b) exhibits obligatory nonfinite participial inflection on the main verb (*close*), which necessitates the auxiliary verb BE (*was*) in finite constructions. Active constructions such as (1b) dictate that the subject proceed, and the object follow, a transitive verb (such as CLOSE), owing to the canonical SVO word order. In the corresponding passive form (2b), the patient-like semantic argument (*a book*) appears in subject position, mirroring the Croatian data (1-2) in that KNJIGA bears accusative object-marking inflection (*knjigu*) in the active (1), and nominative subject-marking inflection (*knjiga*) in the passive (2). Additionally, the passive constructions of both languages (2 and 2b) permit the optional reintroduction of the agent-like argument via a prepositional phrase.

Certain English transitive verbs, including CLOSE, license passive-like intransitive constructions in which the subject is assigned the semantic role corresponding to the canonical transitive object, as in (3b). These *middle* constructions share several key characteristics with the Croatian impersonal (3), in both the syntactic and semantic components. Syntactically, neither construction permits reintroduction of the agent-like argument, which contrasts with the optional prepositional phrases in (2 and 2b). This fact is fundamentally intertwined with their semantic interpretations; as in the Croatian impersonal (3), the English middle (3b) carries a *less-agentive* reading. This is expanded upon further in later sections, but this contrast is clear when the following two sentences are compared:

a. The boat sank.
b. The boat was sunk.

Imagine hearing these two sentences without any prior context-- they will obviously provoke questions. Upon hearing (a), the question immediately raised is something along the lines of "*How* did it sink?". On the other hand, (b) seems to prompt one to ask "*Who* sank it?". I hypothesize that this disparity stems from the *less-agentive* interpretation inherent in (a), which is somewhat analogous to the reading of the Croatian impersonal (3) described in section (1.1)-- it logically follows from this semantic interpretation that an agent-like argument is not permitted to be reintroduced in constructions such as (3) or (3b).

English additionally has a deverbal adjectival construction (4b), which is similar to the passive (2b) in several ways; the main verb has participial morphology (*closed*) and is accompanied by the auxiliary verb BE. As in the passive, the patient-like argument appears in subject position, but, unlike the passive construction, the reintroduction of the agent-like argument via *by*-phrase is not permitted in the deverbal adjectival form. This is the key relevant difference between the English deverbal adjectival and its Croatian counterpart (4); Croatian permits an agentive argument that is introduced obliquely through a prepositional *od*-phrase.

The *Nondeterministically Enumerated Categorial Grammar* (NEGC) framework-- which is defined and explained in detail throughout the remainder of my paper-- may offer the theoretical foundation necessary to support an analysis of these three Croatian (2-4) and English (y-z') passive-like constructions that is both empirically accurate and explanatorily sound. The following section will first briefly summarize this framework's foundational elements and core distinguishing features. Each mechanism is then illustrated in explicit detail, as all of of the Croatian examples laid out in (1-4) are derived-- each one alongside its corresponding near-synonymous English construction (1b-4b)-- in a step-by-step manner.

(2) An Introduction to NECG

NECG is a curryesque framework, meaning that a formal distinction is maintained between abstract *tectogrammatical* and concrete *phenogrammatical* syntax (Curry 1961), where the *tectogrammatical* sector utilizes *nondirectional* implication of the form $X_1 \Rightarrow \ldots \Rightarrow X_n$ to determine both the order with which a given *functor* term must combine with its arguments in order to reduce down to its *terminal reducing type* (the final *atomic* type X_T to which a given term will eventually reduce), and the requisite (sub-)typing categorization of each argument.

While the *tectogrammar* performs the role of argument selection with respect to *syntactic type*-- selecting only for those terms whose types satisfy its implicative specifications, it is but one of three grammatical components that compose a given valid *term*; each of which is a triple of the form φ; Σ; X where φ represents the *phenogrammatical* component, Σ represents its *semantic* formula, and X, the *tectogrammar*. In contrast to the *contracting* nature of the *tectogrammatical* component, which sequentially eliminates each implication as it is satisfied, eventually reducing to an *atomic/terminal* type, the *phenogrammatical* and *semantic* elements can be thought of as *expanding*; each of these two elements will construct phonological strings and corresponding logical formulae of a predicate calculus over the course of a derivational proof.

The *semantic* sector of any given term consists of either a valid logical formula, or a valid term of the typed λ-calculus defined over the formal predicate logic. As briefly mentioned in the previous section, a one-to-one correspondence (*isomorphism*) can be established for any given term between the order of each λ-abstraction operator binding its semantic formula, and the order of each syntactic argument selected by its *tectogrammatical* implicational formula. In other words, for any number *n* of syntactic arguments selected by a given term's type, *n* respectively corresponding λ-abstraction operators can be introduced to bind variables in the term's *semantic* sector.

As in the *semantic* sector, the *phenogrammar* consists of a typed λ-calculus. In this component, however, the λ-calculus is defined over the monoid $\langle \boldsymbol{\phi}, \bullet, \varepsilon \rangle$ -- the *phonological alphabet* $\boldsymbol{\phi}$, the *string concatenation* binary operator •, and the monoid *identity element* ε (i.e. the empty string). This novel technique of phonological λ-abstraction was pioneered in Oehrle (1994), and further developed in later works e.g. de Groote (2001).

These φ; Σ; X triples are combined to yield grammatical constructions in a fragment of the Natural Deduction proof calculus that-- at least for this analysis-- only utilizes a single (nondirectional) Rule of Inference; *Implication Elimination* (defined in 6 below), and consequently need not introduce or eliminate variables in the course of a derivational proof-- even, as demonstrated in the following sections, to derive scopal ambiguities in a given construction's final semantic interpretation.

6. (*Implication Elimination*)

$$\frac{\varphi_i;\ \Sigma_i;\ X_j \Rightarrow X_i \qquad\qquad \varphi_j;\ \Sigma_j;\ X_j}{\varphi_i(\varphi_j);\ \Sigma_i(\Sigma_j);\ X_i}\ {\scriptstyle \Rightarrow \text{Elim.}}$$

In addition to the *isomorphism* maintained (to a degree-- see section 4.1) between *phenogrammatical/semantic* λ-abstraction and *tectogrammatical* implication, every term exists in a normalized form such that there is a direct correspondence between the left-to-right linear order of a given term's φ-variables the ordering of the phenogrammatical λ-abstraction on these φ-variables. For example, the phenotype for an English ditransitive verb (e.g. *give*) might look something like this; $\lambda\varphi_1\lambda\varphi_2\lambda\varphi_3.[\varphi_1\bullet\text{give}\bullet\varphi_2\bullet\varphi_3]$, while six phonologically identical forms exist for each possible scopal ordering of the arguments. This is possible due to a foundational tenet of this formalism; *the set of terms is disjoint from the set of lexical items*. Each lexical item, rather than existing directly as an element of a set of proof terms, instead consists of an algorithm to nondeterministically enumerate its (possibly transfinite) set of associated phrasal structures with $n \geq 0$ arguments (and $n!$ phonologically identical forms for each possible scopal ordering for each phrasal structure) and is interpreted by the *Pre-Syntactic Automaton*, which recursively enumerates the set of proof terms from the lexicon, and additionally ensures the strict left-to-right correspondence between phonological variables and the scopal ordering of their binding λ-abstraction operators.

The grammar is built around *lexical* terms (i.e. those represented semantically by elements of the underlying logic's signature, as opposed to logical constants e.g. modals, polarity items, quantifiers, etc.), which form the basis for the *complex types*. These are those terms whose tectogrammatical sectors generally consist of implicational types, as opposed to the *atomic types*, whose tectogrammatical types are in irreducible forms (e.g. S, NP, etc.).

This next section will introduce the system and explain the above two paragraphs by deriving the Croatian active voice construction "*Marko je zatvorio knjigu*" and its English equivalent, "*Bob closed a book*".

(2.1) Active Voice

Each lexical entry $L_i \in L$ (where L is the lexicon) is represented as a 2-tuple $<\beta_i, M_i>$, where each $\beta_i: \mu \mapsto \Phi$ is a string to string partial function representing a mapping from a finite subset of the language μ^*, where each letter of μ corresponds to a morphosyntactic value, to a finite subset of Φ^*, the free monoid on the *phonological alphabet*. These are called *paradigm functions*-- not to be confused with those of the *Paradigm Function Morphology* framework (Stewart and Stump 2007)-- and constitute the interface between the syntax, morphology, and phonology, where sequences of abstract morphosyntactic features are composed during the course of a derivation, then translated via β-function into more concrete phonological strings.

To that end, each lexical entry corresponds to a unique, unary *paradigm function* that takes as its argument a finite string of arbitrary length, representing a sequence of morphosyntactic feature values. It then returns the phonological string that corresponds to that lexical entry's morphophonological realization of the concatenated order of the morphosyntactic feature values written on the input string. A subset of the morphosyntactic alphabet (μ) is given below, where the Leipzig Gloss abbreviation of each letter is given on the right side. A given (Latin) letter with a superscripted string of arbitrary length is considered a single character of the alphabet μ (e.g. $P^{ROG} \in \mu$):

7.

1 = 1 (*first person*)
2 = 2 (*second person*)
3 = 3 (*third person*)
SG = S^{G} (*singular*)
PL = P^{L} (*plural*)
PRS = P^{RS} (*present*)
PST = P^{ST} (*past*)
PCTC = P^{RT} (*past participle*)
PROG = P^{ROG} (*present progressive participle*)
ADJ = A^{DJ} (*deverbal adjective*)
NMLZ = N^{MLZ} (*nominalized verb*)
ø = ε (*uninflected form*)

A given *paradigm function*, e.g. that of the lexeme CLOSE-- β_{CLOSE}-- takes a string argument corresponding to a sequence of morphosyntactic feature values, then returns a phonological string, as in the following examples;

8.

$\beta_{CLOSE}(P^{RS}\bullet 3\bullet S^{G})$ = "*closes*" (*3SG*)
$\beta_{CLOSE}(N^{MLZ}\bullet S^{G})$ = "*closing*" (*nominalized verb*)
$\beta_{CLOSE}(N^{MLZ}\bullet P^{L})$ = "*closings*"[1] (*nominal & plural*)
$\beta_{CLOSE}(P^{RT})$ = "*closed*" (*past participle*)
$\beta_{TRANSPORT}(N^{MLZ}\bullet S^{G})$ = "*transportation*"

[1]: e.g. "Many banks on Wall Street experienced closings after the recession."

However, a *paradigm function*, is only a *partial* function-- the domain of all β_i functions is that finite subset of the free monoid on the alphabet μ corresponding to *valid sequences of morphosyntactic feature values*-- no *paradigm function* has a defined output corresponding to an invalid input string e.g. $P^{RS}\bullet S^{G}\bullet S^{G}$.

None of which, of course, explains the mechanism by which these morphosyntactic strings are composed and passed to a given β_i function. As it so happens, this mechanism happens to be a component of the same system that enumerates the set of φ; Σ; X triples which comprise the valid terms of the proof calculus; *the Pre-Syntactic Automaton*. The reasoning behind this is based on a fairly uncontroversial premise; for any given lexeme, there is direct correspondence between its overt phonological realization and syntactic combinatorics. In English, for example, nonfinite verbs do not permit nominative pronominal subjects, singular nouns require a determiner to form an NP (disregarding mass and proper nouns of course), and so on. Straightforward examples from other languages include applicative affixes (Austin 2005) and other valence-changing morphological processes. Instead of deriving syntactic combinatorics from the morphosyntactic feature specifications corresponding to phonological strings or vice versa, the NECG framework composes both in parallel. The rest of this section-- a derivation of the English active sentence "*Bob closed a book*"-- contains a step-by-step example of this process.

As previously described, each lexeme corresponds to a set of one or more proof terms of the form $\lambda\varphi_1 ... \lambda\varphi_n.[\varphi_1 \bullet \ldots \bullet \varphi_n]; \lambda\mathcal{P}_1 ... \lambda\mathcal{P}_n.[\Sigma]; X_1 \Rightarrow \ldots \Rightarrow X_n \Rightarrow X_T$, where each φ_i is a λ-bound phonological string variable, each $\mathcal{P}_i$ is a polymorphic variable over any given semantic type, and each X_i is a given tectogrammatical type, while X_T is the term's *terminal reducing type*. In the derivational proof, the final term will be reduced down to a triple of the form φ; Σ; X, representing a phonological string, its corresponding semantic interpretation, and its tectogrammatical type.

As in section (1.1), the set of types consists of the following two subsets; *atomic types* and *complex types*, where the set of *atomic types* = {NP, AdjP, AdvP, Det, PP, S} (with subtyping for any given atomic type e.g. S_{FIN}, S_{INF}, etc.). The *complex types* are defined as follows; if A and B are *types*, A ⇒ B is a *complex type*. As an example, one of the terms corresponding to the lexeme CLOSE is given in (9);

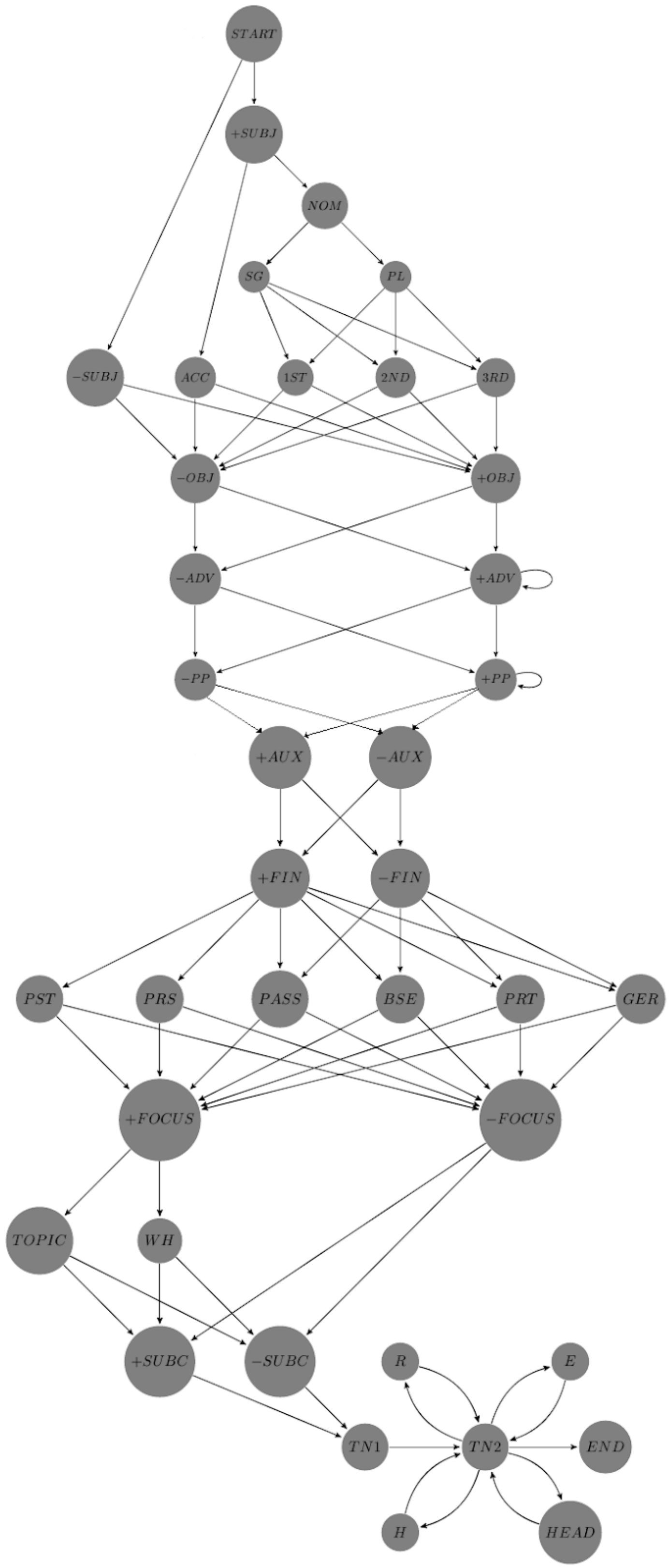
9a. $\lambda\varphi_1\lambda\varphi_2.[\varphi_1\bullet\text{closed}\bullet\varphi_2]; \lambda\mathcal{P}_1\lambda\mathcal{P}_2.\mathcal{P}_1(\lambda x.\mathcal{P}_2(\lambda y.[\mathit{close}(\mathit{arg}_1(x) \cap \mathit{arg}_2(y))])); NP_{NOM} \Rightarrow NP_{ACC} \Rightarrow S$

This will derive the English active sentence "*Bob closed a door*", when combined with two terms of the type NP_{NOM} and NP_{ACC}, corresponding to the strings *Bob* and *a door*, respectively. However, the same CLOSE lexeme also corresponds to a term e.g.;

9b. $\lambda\varphi_1\lambda\varphi_2\lambda\varphi_3.[\varphi_1\bullet\text{closed}\bullet\varphi_2\bullet\varphi_3];$
$\lambda\mathcal{P}_1\lambda\mathcal{P}_2\lambda\mathcal{P}_3.\mathcal{P}_3(\lambda R.\mathcal{P}_1(\lambda x.\mathcal{P}_2(\lambda y.[\mathit{close}(\mathit{arg}_1(x) \cap \mathit{arg}_2(y) \cap R)])));$
$NP_{NOM} \Rightarrow NP_{ACC} \Rightarrow PP \Rightarrow S$

Which will derive an active sentence with a subject, object, and prepositional phrase e.g. "*Bob closed a door with his foot*". Recall that the set of proof terms is *not* the set of lexical entries, but rather is *enumerated by* the set of lexical entries. To that end, each lexical entry is a directed graph, such as the following, CLOSE;

Figure 2


START
+SUBJ
NOM
SG
PL
−SUBJ
ACC
1ST
2ND
3RD
−OBJ
+OBJ
−ADV
+ADV
−PP
+PP
+AUX
−AUX
+FIN
−FIN
PST
PRS
PASS
BSE
PRT
GER
+FOCUS
−FOCUS
TOPIC
WH
+SUBC
−SUBC
R
E
TN1
TN2
END
H
HEAD

The digraph in figure 2 is essentially an algorithm that will be passed as an input to the *Pre-Syntactic Automaton*, which enumerates the set of possible proof terms derivable from this graph. The automaton begins at the START node, and proceeds to a given node N to which it is connected by an edge of the form <START, N>. In the figure above, the next two possible moves are +SUBJ and -SUBJ. Each node consists of an *Instruction Function*, which performs set theoretic operations on a *global variable* set, and a *Condition Function*, which will return *true* or *false*, depending on the configuration of the *global variable* set. From a given node X, the automaton may proceed to any node Y to which X is connected by an edge of the form <X, Y>, *and* the *Condition Function* of Y returns *true*.

Representing each lexeme in this fashion associates each syntactic combinatorial configuration and morphophonological form with a set of requisite variables; effectively allowing certain phrasal structures corresponding to a given lexeme to license (and forbid) certain phonological realizations of said lexeme and vice versa.

For example, in order to derive the proof term given in (9a) that is required to derive "*Bob closed the door*", CLOSE is first passed to the *Pre-Syntactic Automaton, α*; *α*(CLOSE);

10a.

$\alpha(L_i) =_{def} t_F \in stage_1(L_i)(\{u', y, a_L, a_H, a_R, a_E, a_P, \varphi_{THIS}, X_T, \varphi_T, \Sigma_T\}) \mid y = S_i \in L_i$

This is a function which takes in a given lexeme as its input, and returns a derived proof term as its output. $\{u', y, a_L, a_H, a_R, a_E, a_P, \varphi_{THIS}, X_T, \varphi_T, \Sigma_T\}$ is the set of *initial global variables*, which will be manipulated according to the specifications of the CLOSE graph, and eventually determine the final form of the automaton's output. u' is an ordered n-tuple of $<\Phi, \Sigma, X>$ triples that determines the final overt linear order, and y is the *current node* on the directed graph (y will always begin as the START node). $a_L, a_H \subseteq \mathbf{N}$ contain indices pointing to the elements of u' that contain semantically Lower- and Higher-Order variables, respectively, while $a_R \cup a_E \cup a_P = a_L$, and are the set of indices pointing to entities, and functional and propositional predicate functions, respectively. φ_{THIS} is the output of the lexeme's β_i-function, and φ_T, Σ_T, X_T are the *terminal* (i.e. final) forms of the resulting term's phenogrammatical, semantic, and tectogrammatical sectors, respectively. To begin, CLOSE and the *initial global variables* are passed to the *stage₁* function;

10b

$stage_1(L_i)(Z) =_{def} Z' \mid Q = \{\forall.v_j \in V_i \mid (<y, v_j> \in E_i) \wedge v_{j,1}(Z)\}\ \&\ V_i, E_i \in L_i$

if $Q = ø$: $Z' = stage_2(Z)$

else: $Z = y_2(Z)$, $y \in Z = pop(Q)_1$, $Z' = stage_1(L_i)(Z)$

This function compiles a set Q, consisting of the next possible moves the automation can make, where V_i and E_i are the sets of vertices and edges, respectively, of the graph in question. Each vertex is of the form <C, I>, where C is the *condition function*, and takes the set of *global variables* as its input. It will then return true or false, depending on the configuration of its input. So for $Q = \{\forall.v_j \in V_i \mid (<y, v_j> \in E_i) \wedge v_{j,1}(Z)\}$, Q will be the set of all nodes v_j such that there is an edge $<y, v_j>$ (y is the *current node*), *and* $v_{j,1}(Z) = 1$, where $v_{j,1}$ is the *condition function* of v_j, and Z is the set of *global variables*.

The automaton then applies the *instruction function* of the *current node*, y, to the set of *global variables*, Z. A given node may modify Z in any way, including adding new elements, as long as none of the *initial global variables* (i.e. $\{u', y, a_L, a_H, a_R, a_E, a_P, \varphi_{THIS}, X_T, \varphi_T, \Sigma_T\}$) are deleted. Using the *pop* function, defined in (10c-d), y is then set equal to a random element of Q, and these new values are then passed recursively to the $stage_1$ function.

10c. *pop*(Z) $=_{def}$ <y, Z'> |

$\quad$ if Z = ø: y = ø, Z' = ø

$\quad$ else: $y = z_i \in Z$, $Z' = Z \setminus \{z_i\}$

10d. *pop*(W)(n) $=_{def}$ <y, W'> |

$\quad$ if W = ø: y = ø, W' = ø

$\quad$ else if $W = <w_1,...,w_i>$:

$\quad\quad$ if i = 1: $y = w_1$, W' = ø

$\quad\quad$ else if n = i: $y = w_i$, $W' = <w_1,...,w_{i-1}>$

$\quad\quad$ else if n = 1: $y = w_1$, $W' = <w_2,...,w_i>$

$\quad\quad$ else: $y = w_{\mathbf{n}}$, $W' = <w_1,...,w_{n-1},w_{n+1},...,w_i>$

This function is defined for two input types; unordered sets (Z) and tuples (W). For an unordered set input Z, *pop* will return a 2-tuple <y, Z'>, where y is a random element of Z, and Z' = Z \ {y}. For an n-tuple input, *pop* requires a second argument, n, corresponding to the ordinal number of the element in question. A 2-tuple <y, W'> is then returned, where y is the n^{th} element of W, and W' preserves the order of all elements of W besides y. For example, *pop*(<a, b, c, d, e, f>)(3) = <c, <a, b, d, e, f>>.

Once the set Q = ø (i.e. there are no more possible moves to make), the automaton then passes the set of *global variables* to stage 2, which begins the process of composing a valid proof term from the values of these variables.

For the present derivation, START is the *current node*. START = $<\lambda Z_1.[1]$, $\lambda Z_2.[Z_2 \cup$ {*topic, subj, aux, advleft, advright, obj, pp,* ARG_P, *open_roles*} | *open_roles* = {arg_1, arg_2}]>, where $\lambda Z_1.[1]$ (Z_1 is a variable over unordered sets) is a constant function that will always return 1/*true*, regardless of the content of the set that it takes as its argument (from here on out, $\lambda Z.[1] = CONS^1$). If one is skeptical that a λ-term $\lambda x.[F]$ can bind a sequence that does not contain any instances of the variable x, look no further than the Church numerals-- the number 0 is encoded as $\lambda f \lambda x.x$ (Jansen 2013), i.e. an instruction to carry out a function f zero times. The START node adds eight variables to the *global variable* set, the six n-tuples *topic, subj, aux, advleft, advright, obj,* and *pp*, which all begin with a length of 0, the set ARG_P, which begins as a null set (this will be the argument taken by the proposition *close*), and the two-member *open_roles* set, which contains the semantic roles that have yet to be assigned to a syntactic argument. These will later be concatenated together via the *tuple concatenation operator* $\oplus$: $W \times W \mapsto W$

10e. $W_i \oplus W_j =_{def}$

$$
\begin{aligned}
&\text{if } W_i = \varnothing \land W_j = \varnothing: \varnothing \\
&\text{else if } W_i = \varnothing: W_j \\
&\text{else if } W_j = \varnothing: W_i \\
&\text{else if } W_i = \langle w_1,\ldots,w_n\rangle \land W_j = \langle w'_1,\ldots,w'_m\rangle: \langle w_1,\ldots,w_n,w'_1,\ldots,w'_m\rangle
\end{aligned}
$$

This will then create the tuple u' = (*topic* $\oplus$ *subj* $\oplus$ *aux* $\oplus$ *advleft* $\oplus$ φ_{THIS} $\oplus$ *obj* $\oplus$ *advright* $\oplus$ *pp*) at the end of the algorithm, which will determine the final linear order. *topic* may contain an optional complementizer and likewise optional topicalized material, *subj* is a singleton tuple that contains the subject (if there is one-- a subject is not obligatory in nonfinite constructions), *aux* optionally contains an auxiliary or sequence of auxiliaries (in this system, nonfinite forms can select for auxiliaries, but not vice-versa), and *advleft* contains those adverbs that appear on the left of the verb. *obj* contains the object (this can be a 2-tuple in ditransitive constructions), *advright* the adverbs appearing on the right of the verb, and finally, *pp* contains the prepositional phrases. What is important to keep in mind is that given the right conditions, any of the above tuples can be empty.

The next possible nodes are +SUBJ and -SUBJ, which add or do not add a subject, respectively. Both of these nodes have $CONS^1$ as *condition* functions, so they will both always be possible next moves when START is the current node. In this case, +SUBJ is selected, where +SUBJ = $\langle CONS^1, \lambda Z.[Z \cup \{arg\_subj\} \mid subj \in Z = \langle\langle \varphi_i, x, NP\rangle\rangle \,\&\, arg\_subj = x]\rangle$, which adds a triple $\langle \varphi_i, x, NP\rangle$ to position one of the *subj* tuple, and the semantic entity-type variable x to *arg_subj*, a *global variable* that stores the subject's semantic variable until its role is determined (as the subject of CLOSE can have an agent-like role in active constructions, or a patient-like one in passive or middle constructions e.g. "*the book closed last night*"). -SUBJ = $\langle CONS^1, ID^Z\rangle$, where $ID^Z = \lambda Z.Z$, and is a set-theoretic identity function (so -SUBJ does *not* make any changes to the set of *global variables*). From +SUBJ, the next possible moves are NOM and ACC. NOM is selected, which sets $subj_{1,3}$ (the third element of the first element, i.e. *NP* in the triple $\langle \varphi_i, x, NP\rangle$-- the first and only element of *subj*) equal to NP_{NOM}. The next possible values are the SG and PL agreement values, which are both connected via edges to 1ST, 2ND, and 3RD. These five *agreement nodes* set the person and number subtype values for the subject NP, and determine the agreement inflection for CLOSE. SG and 3RD are selected in this scenario, and SG = $<CONS^1, \lambda Z.[Z \cup \{agr\_num\} \mid subj_{1,3} \in Z = NP_{NOM,SG}\ \&\ agr\_num = SG]>$, which sets $subj_{1,3}$ (i.e. the subject's tectogrammatical type) to $NP_{NOM,SG}$ and adds the *global variable agr_num* (*agr_num* = SG), which will later be one of the variables that determine the input of the β_{CLOSE} *paradigm function*. 3RD = $<CONS^1, \lambda Z.[Z \cup \{agr\_per\} \mid subj_{1,3} \in Z = NP_{NOM,agr\_num,3}\ \&\ agr\_per = 3]>$, which performs a similar role as SG, but sets the subject person agreement to *third person*, rather than determining the number agreement value. In this derivation, the tectogrammatical type of the subject NP changes from $NP_{NOM,SG}$ to $NP_{NOM,SG,3}$ at this node.

From the person agreement values, the automaton can then select +OBJ and -OBJ, which perform similar functions as their +SUBJ and -SUBJ counterparts, but instead determine the presence, or lack thereof (respectively), of an object argument. +OBJ = $<CONS^1, \lambda Z.[Z \mid obj \in Z = <<\varphi_j, y, NP_{ACC}>>\ \&\ ARG\_P = arg_2(y)\ \&\ open\_roles = \{arg_1\}]>$. Note that +OBJ does not add a *global variable* representing its semantic variable, as, unlike the subject, the object of CLOSE can only be assigned a patient-like semantic role, regardless of any other factors. Additionally, if CLOSE takes a (syntactic) object argument, this implies that the subject (if there is one) must be assigned the agent-like/arg_1 role-- neither of the two constructions that assign a patient-like role to the subject (namely, the passive and middle) permit an object argument. As such +OBJ sets the *open_roles* set to $\{arg_1\}$, limiting the subject argument's potential role(s) to the arg_1 semantic role.

At some point in this section, the inquisitive reader may have noticed something peculiar about the semantic function of the term given for *close*; $\lambda\mathcal{P}_1\lambda\mathcal{P}_2.\mathcal{P}_1(\lambda x.\mathcal{P}_2(\lambda y.[close(arg_1(x) \cap arg_2(y))]))$ namely, *close* is a unary propositional predicate (i.e. *close*: $Z \mapsto \{1, 0\}$) containing two functional predicates (i.e. arg_1, arg_2: $Z \mapsto Z$), arg_1 and arg_2, both of which are also unary. These two functions, $arg_1(x)$ and $arg_2(y)$, return the subset of *relations* for which x is the *primary argument* and y is the *secondary argument*, respectively. *close*(X) then returns true if an element of this set is a subset of an element of the union of all ***close*** relations, and false otherwise. This represents a variant of First-Order monadic logic (Frick and Grohe 2004) that permits unary functional predicates, and the binary *set intersection* operator.

For example, the sentence [*John*] *closed* [*War and Peace*] [*on the table*], has three functional predicates (bracketed) in its semantic interpretation; $close(arg_1(john) \cap arg_2(wap) \cap on(\iota(tbl)))$. $arg_1(john)$ returns a set e.g. $\{...,\{<john, arg_1>, <wap, arg_2>, <\iota(tbl), on>\},...,\{<john, arg_1>, <wap, arg_2>, <mary, with>\},...,\{<john, arg_1>, <sue, arg_2>\},...\}$ -- this is the set of all *relations* in which *John* has the arg_1 role. This is then intersected with the output of $arg_2(wap)$-- all of the relations for which *War and Peace* is the *secondary argument*-- to yield the following; $\{...,\{<john, arg_1>, <wap, arg_2>, <\iota(tbl), on>\},...,\{<john, arg_1>, <wap, arg_2>, <mary, with>\},...\}$, which is again intersected with the set of relations "occurring on the table"; $on(\iota(tbl)))$ to yield; $\{...,\{<john, arg_1>, <wap, arg_2>, <\iota(tbl), on>\},...\}$.

close(X) then checks that $\exists y \exists z[(y \in \mathbf{close}) \wedge (z \in X) \wedge (z \subseteq y)]$, so if there exists a set in the set of **close** relations such that $\{<john, arg_1>, <wap, arg_2>, <\iota(tbl), on>\}$ is a subset or equal to of that set, $close(arg_1(john) \cap arg_2(wap) \cap on(\iota(tbl)))$ will return 1, and 0 otherwise. This allows one to interpret constructions such as "*a ball was given yesterday*", in which a semantically ternary predicate takes only one syntactic argument, without having to make claims about the (non-)existence of the other two arguments in the predicate e.g. $\exists(\mathbf{ball})(\lambda u.\exists(\mathbf{x})(\lambda v.\exists(\mathbf{y})(\lambda w.[give(v)(u)(w)])))$ through e.g. existential quantification. In this system, that same predicate can be represented as $\exists(\mathbf{ball})(\lambda x.[give(arg_2(x))])$.

Returning to the pre-syntactic term derivation, +OBJ adds a triple $<\varphi_j, y, NP_{ACC}>$ to the first element of the (singleton) *obj* tuple, and additionally sets ARG_P equal to $arg_2(y)$.

+ADV and -ADV allow the system to optionally (viz. nondeterministically) add adverbs, and there is an edge connecting +ADV to itself, allowing the possibility of recursive addition of adverbs (theoretically) ad infinitum. The nodes +PP and -PP are very similar to +ADV and -ADV, respectively, but add (or, in the case of -PP, chose not to add any) PP-type arguments. +AUX and -AUX add an auxiliary verb, or don't, respectively. +PP and +AUX will be defined and described in later sections, but in the current derivation, the system chooses the path -ADV ⟶ -PP ⟶ -AUX. -ADV and -PP = $<CONS^1, ID^Z>$, which means that they have no *conditions* (i.e. are always an available path for the automaton to take) and make no changes to the set of *global variables*. -AUX = $<CONS^1, \lambda Z.[Z \cup \{auxiliary\} \mid auxiliary = 0]>$, and adds a boolean (i.e. valued either 1/*true* or 0/*false*) *global variable*, *auxiliary*, which is subsequently set to 0 (*false*).

+FIN and -FIN, in a similar fashion to +AUX and -AUX, introduce a boolean *global variable*, *finite*, and set its value to 1 (+FIN) and 0 (-FIN). +FIN = $<\lambda Z_1.[(subj \in Z_1) \wedge subj_{1,3} = NP_{NOM}], \lambda Z_2.[Z_2 \cup \{finite\} \mid finite = 1 \ \& \ ARG\_P \in Z = ARG\_P \cap \{pop(open\_roles)_1(arg\_subj)\}]>$, which states that it is only an accessible node if $subj_{1,3}$ (i.e. the subject's tectogrammatical type) is a nominative NP, and sets the value of the boolean *finite* to 1. In addition, the *instruction function* sets ARG_P equal to itself intersected with a singleton set containing a semantic role that is nondeterministically selected from the set of available semantic roles (*open_roles*), such that the subject NP's semantic variable, *arg_subj*, is taken as said semantic role's argument. As there is an object present in the current derivation, *open_roles* = $\{arg_1\}$, and so, in this scenario, ARG_P = ARG_P ∩ $\{arg_1(arg\_subj)\}$ is the only possible outcome.

However, if the -OBJ node were instead to have been selected, *open_roles* = {arg_1, arg_2}, as only the +OBJ node removes arg_2 from the *open_roles* set. In this hypothetical scenario, arg_2 could potentially be selected as the subject's semantic role; this is how the middle voice construction, e.g. "*the book closed last night*", is obtained. However, in the current derivation, there does exist an object argument, and as such only the active construction can be derived.

The *auxiliary* and *finite* booleans, instantiated by the +FIN/-FIN and +AUX/-AUX nodes, respectively, are used to determine whether the construction takes one of four forms regarding to the (main) verb's finiteness and the presence (or lack thereof) and finiteness of an auxiliary verb, as schematized in the following table:

	auxiliary = 1	*auxiliary* = 0
finite = 1	[NFT verb]&[FIN aux]	[FIN verb]&[¬∃.aux]
finite = 0	[NFT verb]&[NFT aux]	[NFT verb]&[¬∃.aux]

In this case, "*closed*" (*past tense*, rather than *past participle*) is a finite base verb form, and so can only be derived when *auxiliary* = 0 and *finite* = 1. Note that PRS and PST are connected via edges to the +FIN node only, meaning that the automaton *cannot* move to these nodes from -FIN, regardless of their *condition function* outputs. In this scenario, the system advances to the PST (past tense) node. PST = <$\lambda Z_1.[(\mathit{finite} \in Z_1) \wedge (\mathit{auxiliary} \in Z_1) \wedge (\mathit{finite} \wedge \neg \mathit{auxiliary})], \lambda Z_2.[Z_2 \mid X_T \in Z_2 = S_{PST}$ & $\varphi_{THIS} \in Z_2 = \beta_{CLOSE}(P^{ST} \bullet \mathit{agr\_per} \bullet \mathit{agr\_num})]$> -- the *condition function* of PST returns 1 iff *finite* = 1 and *auxiliary* = 0, and its *instruction function* sets the value of the *terminal reducing type* variable, X_T, to S_{PST} (S_{PST} is a subtype of S_{FIN}) and φ_{THIS} (essentially synonymous with a *phrase head*-- the phonological string representing a given lexeme) is set equal to the output of $\beta_{CLOSE}(P^{ST} \bullet 3 \bullet S^G)$--"*closed*"-- as *agr_per* = 3 and *agr_num* = S^G.

From PST, the system has the option to proceed to either +FOCUS or -FOCUS (+FOCUS allows for *wh*- and topicalization focus), and, as the term that is currently being derived does not have a topicalized/focused element, -FOCUS is selected in the course of this derivation. -FOCUS = <$CONS^1$, ID^Z> -- in other words, it has a constant (i.e. always 1/*true*) *condition function*, and makes no changes to the set of *global variables* if selected. From -FOCUS, the system may proceed to either +SUBC or -SUBC, which, as the names imply, designate the currently deriving term as a subordinate clause, or not, respectively. As the construction in question ("*Bob closed the book*") is not a subordinate clause, the system will proceed to -SUBC, which has the same form (and consequently, function-- i.e. "do nothing") as -FOCUS; -SUBC = <$CONS^1$, ID^Z>.

The system will then proceed to TN1 ("Terminal Node 1"), where TN1 = <$CONS^1$,

$\lambda Z.[Z \cup \{length, count\} \mid u' \in Z = (topic \oplus subj \oplus aux \oplus advleft \oplus \varphi_{THIS} \oplus obj \oplus advright \oplus pp)$ & $length = length'(u')$ & $count = 0$>, which sets u' equal to the concatenation of the six *phrase-positional* n-tuples, along with the *phrase head*, φ_{THIS}, and additionally adds two nonnegative integers to the set of *global variables*, *length* and *count*, where *count* = 0 and *length* is set equal to the new length of u'. In this case, u' = <<φ_1, x, NP_{NOM}>, *closed*, <φ_2, y, NP_{ACC}>>, and as such, *length* = 3. The *length* and *count* variables are used to assist in a recursive process that will cycle through each argument (i.e. all elements of u' that are not φ_{THIS}) individually, and place them into one of three sets, a_H, a_E, and a_R (denoting *Higher-Order*, *entity*, and *functional predicate* semantic types, respectively), all of which will be interpreted by later stages of the *Pre-Syntactic Automaton* to compile the construction's final semantic λ-term.

With that goal, the system then proceeds to TN2 ("Terminal Node 2"), which has the following form; TN2 = <$CONS^1$, $\lambda Z.[Z \mid count \in Z = count + 1]$>, which unconditionally increments the *count* variable by one. At this point in the pre-syntactic derivation, *count* = 1, as the system must necessarily move from TN1 to the only connected node; TN2. There are now five nodes to which TN2 is connected via an edge; H, R, E, HEAD, and END. However, the system is constructed in such a way that only one of the five's *condition functions* will return 1 at a time, and as such there will always *exactly one* move that the automaton is permitted to make-- from this point in the derivation until it reaches the *terminal* END node, and stage 1 of the *Pre-Syntactic Automaton*'s derivation is complete.

E = <$\lambda Z_1.[u'_{count,3} \in Z_1 = NP]$, $\lambda Z_2.[Z_2 \mid a_E \in Z_2 = a_E \cup \{count\}]$>, and its *condition function* returns 1 iff the third element of the n^{th} element of u' such that *n* = *count* is of type NP. In other words, if the tectogrammatical type of u'_n = NP-- each element of u', with the exception of φ_{THIS}, is of the form <φ, Σ, X>. This makes the fairly reasonable assumption that all NPs are semantically represented by entity-types, following the example of many forms of Type-Logical Grammar (Moortgat 2010). The *instruction* of E then adds value of *count* (*count* ∈ **N**) to the set a_E-- this integer is an index pointing to u'_n, the NP/entity-type element in question.

R = <$\lambda Z_1.[u'_{count,3} \in Z_1 = PP \lor u'_{count,3} = AdvP]$, $\lambda Z_2.[Z_2 \mid a_R \in Z_2 = a_R \cup \{count\}]$>, and adds natural number indices to the set a_R, such that each index points to a unique element of the n-tuple u' whose tectogrammatical type is either PP or AdvP (the two types whose semantic values are of the *functional predicate* Σ-type, and can also be taken as verbal arguments).

HEAD = <$\lambda Z_1.[u'_{count} \in Z_1 = \varphi_{THIS}]$, ID^Z>, which essentially serves as a "discharge" for the inevitable moment when $u'_{count} = \varphi_{THIS}$ (the *phrase head*), and makes no change to the set of *global variables*, as φ_{THIS} is a member of u' purely for purposes of linearization with respect to its arguments.

Finally, H = <$\lambda Z_1.[u'_{count,3} \in Z_1 \neq PP \land u'_{count,3} \neq AdvP \land u'_{count,3} \neq NP \land u'_{count} \neq \varphi_{THIS}]$, $\lambda Z_2.[Z_2 \mid a_H \in Z_2 = a_H \cup \{count\}]$>, and operates off of the assumption that anything not falling into the tectogrammatical categories accounted for in the E or R nodes must be a Higher-Order type (e.g. modal operators, polarity items, etc.). This node adds a natural number index to a_H for every n^{th} element of u' such that $n = count$, and H is the current node to which the automaton has moved when $n = count$. In other words, H will add a respective index to a_H for every element of u' that is not of the *functional predicate* or *entity* Σ-types.

While the set a_P is reserved for arguments of the *propositional predicate* Σ-type(i.e. *truth-conditional*/$e \rightarrow t$), CLOSE cannot take a subordinate clause as an argument, and as such should never have to handle elements of u' of that type. With the exception of END, four of these five nodes (E, R, H, and HEAD), are connected to TN2 (and *only* TN2) bidirectionally, and will always proceed back to that node in the following move, where TN2 will then increment *count* by one, and return to one of those four nodes, until *count* > *length*, at which point the only possible move is to the *terminal* END node.

Returning to the derivation, which is currently at node TN1, the system will then perform the following moves (the superscripted numbers represent the value of *count* at a given node); $TN1^0 \rightarrow TN2^1 \rightarrow E^1 \rightarrow TN2^2 \rightarrow HEAD^2 \rightarrow TN2^3 \rightarrow E^3 \rightarrow TN2^4$, which will add the indices 1 and 3 to a_E-- 1 and 3 point to the subject and object NPs, respectively. At this point, *count* = 4, satisfying the *condition function* for the *terminal node*, END.
END = $<\lambda Z_1.[count \in Z_1 > length \in Z_1], \lambda Z_2.[Z_2 \mid \Sigma_T \in Z_2 = close(\text{ARG\_P} \in Z_2)]>$ -- this *condition function* states that the system is permitted to move to END iff the value of *count* is greater than the length of u' (i.e. the variable *length*), and sets the *terminal semantic type* Σ_T equal to *close*(ARG_P). In this case, ARG_P = $arg_1(x) \cap arg_2(y)$, so $\Sigma_T = close(arg_1(x) \cap arg_2(y))$.

From the node END, the set Q of possible nodes will be ø, and as such this final set of *global variables* will then be passed to the $stage_2$ function, which begins the process of constructing a valid proof term; a triple of the form φ; Σ; X.

10f. $stage_2(Z) =_{def} Z' \mid$
if $(a_H \cup a_L = ø)$: $Z' = stage_5(Z)$
else if $(a_L \neq ø \land pop(\{1, 0\})_1)$: $Z' = stage_3(Z)$
else if $a_H \neq ø$: $Z' = stage_4(Z)$
else: $Z' = stage_2(Z)$

10g. $stage_3(Z) =_{def} stage_2(Z') \mid <A, <\Sigma_i, n>> = pop(a_L \in Z), Z' = Z, a_L \in Z' = A,$
$\Sigma_T \in Z' = \mathcal{P}_k(\lambda\text{-abs}(\Sigma_i)(\Sigma_T)), u'_{n,2} = \mathcal{P}_k$

10h.
$stage_4(Z) =_{def} stage_2(Z') \mid <A, <\mathcal{P}_i, n>> = pop(a_H \in Z), Z' = Z, a_H \in Z' = A, \Sigma_T \in Z' = \mathcal{P}_i(\Sigma_T)$

Stage 2 applies recursively until $a_H \cup a_L = ø$. It nondeterministically proceeds to stages 3 (10g) or 4 (10h), corresponding to Lower- and Higher-Order semantic types, respectively. Stage 3 pulls a random element e from a_L, and then $a_L = a_L \setminus \{e\}$. e is then λ-abstracted, and binds its corresponding variable within Σ_T using the *λ-abs* function:

10i. $\lambda\text{-}abs(\mathcal{P}_i)(\Sigma_k) =_{def}$
$\lambda\mathcal{P}_i.[\Sigma_k]$

The triple $<\varphi_e, \Sigma_e, X_e>$ corresponding to e is then set equal to $<\varphi_e, \mathcal{P}_i, X_e>$, while $\Sigma_T = \mathcal{P}_i(\lambda\text{-}abs(\Sigma_e)(\Sigma_T))$. While it isn't used here, stage 4 (10h) corresponds to Higher-Order semantic types, e.g. modal operators, negation, etc., and pulls an element e from a_H, $a_H = a_H \setminus \{e\}$ then $\Sigma_T = e_2(\Sigma_T)$, where e_2 is the second element of e, i.e. its semantic variable.

Stages 3 and 4 both obligatorily return to stage 2, and as such stages 2, 3, and 4 together comprise a sort of "composite" function that nondeterministically and recursively cycles through its three sub-functions until the sets a_L and a_H are both empty. In this case, the object argument is first pulled (nondeterministically) from a_L, λ-abstracted, and $\Sigma_T = \mathcal{P}_2(\lambda y.[close(arg_1(x) \cap arg_2(y))])$. The subject is then pulled, λ-abstracted, and $\Sigma_T = \mathcal{P}_1(\lambda x.\mathcal{P}_2(\lambda y.[close(arg_1(x) \cap arg_2(y))]))$.

Note that the term $\mathcal{P}_2(\lambda y.\mathcal{P}_1(\lambda x.[close(arg_1(x) \cap arg_2(y))]))$ is also derivable, owing to the nondeterministic nature of the automaton. It is from this mechanism that different scopal orders may be obtained; for any configuration of $\{a_L \cup a_H\}$ a cardinality n, stage 2 will nondeterministically derive n! Σ_T terms, each one corresponding to a different scopal ordering.

At this point in the current derivation, $a_H \cup a_L = ø$, and as such the automaton now passes this newly modified set of *global variables* to stage 5, which is an intermediary stage that simply adds a count variable, i = 1, and a variable j equal to the length of u' (in this case, j = 3) to the set of *global variables*, then proceeds to stage 6;

10j.
$stage_6(Z) =_{def} Z'$ |
if i > j: $Z' = stage_7(Z)$
else if i > j:
if $u'_i = \varphi_{THIS}$: $\varphi_T = \varphi_T \bullet \varphi_{THIS}$, $w' = pop(u')(i)$, $u' = w'_2$, $j = j - 1$, $Z' = stage_6(Z)$
else: $\varphi_T = \varphi_T \bullet u'_{i,1}$, $i = i + 1$, $Z' = stage_6(Z)$

Stage 6 then concatenates all of the φ-variables and the φ_{THIS} string in their linear order as determined by their position in u', such that the first element of u' is the leftmost φ-variable in the linear order, and the last element of u', the rightmost. To this end, stage 6 concatenates $u'_{i,1}$ to φ_T and increments i to i + 1 until i > j. φ_T is the *terminal* φ-string/λ-term that will eventually compose the φ-sector of the proof term outputted at the end of the derivation-- in this case, the final form of φ_T at the end of stage 6 is $\varphi_1 \bullet \text{closed} \bullet \varphi_2$. Once i > j, the set of *global variables* is then passed to stage 7;

10k.

$stage_7(Z) =_{def} Z'$ |

if $u' = ø$: $Z' = stage_8(Z \cup \{t_F\})$ | $t_F = ø$

else: $\langle u, u'' \rangle = pop(u')(|u'|)$, $u' = u''$, $\varphi_T = \lambda\text{-}abs(u_1)(\varphi_T)$, $\Sigma_T = \lambda\text{-}abs(u_2)(\Sigma_T)$,

$X_T = u_3 \Rightarrow X_T$, $Z' = stage_7(Z)$

Yet another recursive function, $stage_7$ removes u'_n from u', where n is equal to the length of u'; this removes the last (i.e. rightmost) element from the tuple u'-- recall that each element of u' is a $\langle \varphi, \Sigma, X \rangle$ triple, and the order imposed on u' is isomorphic to the linear order of the concatenated φ-variables corresponding to each respective element of u' in φ_T (the phonological sector of the final φ; Σ; X proof term). For example, in this current derivation, $u' = \langle\langle \varphi_1, \mathcal{P}_1, NP_{NOM} \rangle, \langle \varphi_2, \mathcal{P}_2, NP_{ACC} \rangle\rangle$, and $\varphi_T = \varphi_1 \bullet \text{closed} \bullet \varphi_2$.

This stage λ-abstracts the removed element's φ- and Σ-sectors, and introduces an implication of the form $X_T = X_i \Rightarrow X_T$, where X_i is the tectogrammatical type of the just-removed (ex-)element of u'. In this case, these the following values for u', φ_T, Σ_T, and X_T following $stage_6$, but before the first application of $stage_7$;

$u' = \langle\langle \varphi_1, \mathcal{P}_1, NP_{NOM} \rangle, \langle \varphi_2, \mathcal{P}_2, NP_{ACC} \rangle\rangle$

$\varphi_T = \varphi_1 \bullet \text{closed} \bullet \varphi_2$

$\Sigma_T = \mathcal{P}_1(\lambda x.\mathcal{P}_2(\lambda y.[close(arg_1(x) \cap arg_2(y))]))$

$X_T = S_{PST}$

Following the first application of $stage_7$, the following values are obtained;

$u' = \langle\langle \varphi_1, \mathcal{P}_1, NP_{NOM} \rangle\rangle$

$\varphi_T = \lambda\varphi_2.[\varphi_1 \bullet \text{closed} \bullet \varphi_2]$

$\Sigma_T = \lambda\mathcal{P}_2.[\mathcal{P}_1(\lambda x.\mathcal{P}_2(\lambda y.[close(arg_1(x) \cap arg_2(y))]))]$

$X_T = NP_{ACC} \Rightarrow S_{PST}$

The second (and, in this derivation, final) application of $stage_7$ will then yield the following;

$u' = \varnothing$
$\varphi_T = \lambda\varphi_1\lambda\varphi_2.[\varphi_1 \bullet \text{closed} \bullet \varphi_2]$
$\Sigma_T = \lambda\mathcal{P}_1\lambda\mathcal{P}_2.[\mathcal{P}_1(\lambda x.\mathcal{P}_2(\lambda y.[\mathit{close}(\mathit{arg}_1(x) \cap \mathit{arg}_2(y))]))]$
$X_T = NP_{NOM} \Rightarrow NP_{ACC} \Rightarrow S_{PST}$

Now that the length of u' is 0, the set of *global variables* is finally passed to $stage_8$, which simply combines the φ_T, Σ_T, and X_T values derived by stage 7 into a single term of the form φ_T; Σ_T; X_T, and is defined as follows;

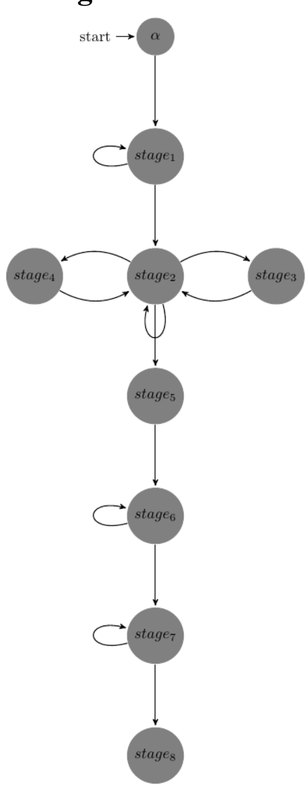
$stage_8(Z) =_{def} Z \mid t_F \in Z = \varphi_T; \Sigma_T; X_T$

As such, the final term necessary to derive an active sentence with a subject and an object, e.g. "*Bob closed a book*", is finally outputted;

$\lambda\varphi_1\lambda\varphi_2.[\varphi_1 \bullet \text{closed} \bullet \varphi_2]; \lambda\mathcal{P}_1\lambda\mathcal{P}_2.\mathcal{P}_1(\lambda x.\mathcal{P}_2(\lambda y.[\mathit{close}(\mathit{arg}_1(x) \cap \mathit{arg}_2(y))])); NP_{NOM} \Rightarrow NP_{ACC} \Rightarrow S$

For visual reference, the directed graph in figure 3 represents the control flow between the eight *Pre-Syntactic Automaton* stages:

Figure 3



The set of *terms* in the proof calculus is then defined as the following:

11. $\forall x[(x \in \textit{Terms}) \leftrightarrow ((x \in A^*) \wedge \exists i[(i \in \mathbf{N}) \wedge (\textit{gdln}(\textit{enc}^A)(x) \in \textit{cycle}(\textit{enc}^A)(i)(L))])]$

Where *gdln* is a function that returns the *Gödel number*[1] corresponding to a given sequence, enc^A is a set of tuples $<a_i, n_i>$ pairing each element of A (where A is the alphabet over all symbols that could potentially be used in a sequence representing a given $<\varphi, \Sigma, X>$ triple) with a unique natural number, and *cycle* returns the set of *Gödel numbers* for a given encoding (enc^A) with a given number of cycles (i) through the *Pre-Syntactic Automaton* for a given lexicon (L). For example, $cycle(enc^A)(5)(L)$ will return a set containing the five *Gödel numbers* representing five random sequences derived by the automaton from the lexicon L.

So the FOL statement in (11) roughly states; "for all x, x is an element of the set of proof calculus terms (*Terms*) if and only if there exists a natural number i such that an element of the output set of $cycle(enc^A)(i)(L)$ is equal to the *Gödel number* of x."

(2.1.1) English Active Voice

At this point in the derivation, the term $\lambda\varphi_1\lambda\varphi_2.[\varphi_1\bullet\text{closed}\bullet\varphi_2]; \lambda\mathcal{P}_1\lambda\mathcal{P}_2.\mathcal{P}_1(\lambda x.\mathcal{P}_2(\lambda y.[\mathit{close}(\mathit{arg}_1(x) \cap \mathit{arg}_2(y))])); NP_{NOM} \Rightarrow NP_{ACC} \Rightarrow S$ has already been obtained via the *Pre-Syntactic Automaton* in the previous section, but "*a book*" and "*Bob*" still must be derived in order to begin the proof. The term for BOB is fairly trivial; bob; $\lambda\mathcal{P}.\mathcal{P}(bob)$; $NP_{NOM,ACC}$, and is a two-node lexical entry, merely to satisfy the requirements that each entry have a distinct *initial* and *terminal* node. As proper names do not inflect, there is only one possible phrasal structure for BOB. Note that the occupant of the semantic sector is the λ-term $\lambda\mathcal{P}.\mathcal{P}(bob)$-- in NECG, each argument is assumed to be n^{th}-Order, and Lower-Order terms are "snuck in" via the n^{th}-Order polymorphic *λ-cloak*; $\lambda\mathcal{P}.\mathcal{P}(x)$. When $\lambda\mathcal{P}_3.\mathcal{P}_3(bob)$ is applied to a term of the form $\lambda\mathcal{P}_1\lambda\mathcal{P}_2.\mathcal{P}_1(\lambda x.\mathcal{P}_2(\lambda y.P'(x)(y)))$, where P' is some arbitrary binary predicate, the following β-reduction steps result (brackets superscripted for legibility):

12. $\lambda\mathcal{P}_1\lambda\mathcal{P}_2.[^1\mathcal{P}_1(\lambda x.\mathcal{P}_2(\lambda y.P'(x)(y)))]^1(\lambda\mathcal{P}_3.[^2\mathcal{P}_3(bob)]^2)$
$\lambda\mathcal{P}_2.[^1\lambda\mathcal{P}_3.[^2\mathcal{P}_3(bob)]^2(\lambda x.[^3\mathcal{P}_2(\lambda y.P'(x)(y))]^3)]^1$
$\lambda\mathcal{P}_2.[^1\lambda x.[^2\mathcal{P}_2(\lambda y.P'(x)(y))]^2(bob)]^1$
$\lambda\mathcal{P}_2.\mathcal{P}_2(\lambda y.[P'(bob)(y)])$

---

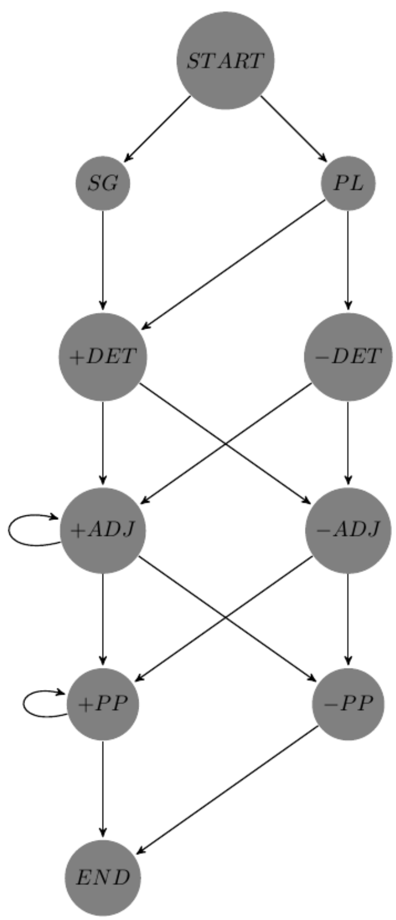
[1] *Gödel numbering*; each element of an alphabet is assigned a unique positive integer, and a sequence of length *k* is then encoded $prime_1^x \times \ldots \times prime_k^y$, where each $prime_i$ is an element in the sequence of prime numbers ($prime_1 = 2$, $prime_2 = 3$, $prime_3 = 5$, $prime_4 = 7$, $prime_5 = 11$, and so on) and the superscripted variables (*x* and *y* in the above example) represent the integers assigned to the character at that position in the sequence. For example, for a three letter alphabet $A = \{a, b, c\}$, $a = 1$, $b = 2$, and $c = 3$. The sequence *abc* is then encoded as $2^1\times3^2\times5^3 = 2250$, $bac = 2^2\times3^1\times5^3 = 1500$, $cba = 2^3\times3^2\times5^1 = 360$, $cbabc = 2^3\times3^2\times5^1\times7^2\times11^3 = 23478840$, etc. What is important is that, owing to the *fundamental theorem of arithmetic*, each unique sequence is paired with a unique *Gödel number*.

Whereas if the argument were quantified, e.g. *a man*--
a•man; $\lambda\mathcal{P}_3.\exists(\mathbf{man})(\lambda z.\mathcal{P}_3(z))$; $NP_{NOM,\ ACC}$, the following β-reduction steps can be obtained (again, brackets are superscripted for legibility):

13. $\lambda\mathcal{P}_1\lambda\mathcal{P}_2.[^1\mathcal{P}_1(\lambda x.\mathcal{P}_2(\lambda y.P'(x)(y)))]^1(\lambda\mathcal{P}_3.[^3\exists(\mathbf{man})(\lambda z.\mathcal{P}_3(z))]^3)$
$\lambda\mathcal{P}_2.[^1\lambda\mathcal{P}_3.[^2\exists(\mathbf{man})(\lambda z.\mathcal{P}_3(z))]^2(\lambda x.\mathcal{P}_2(\lambda y.P'(x)(y)))]^1$
$\lambda\mathcal{P}_2.[^1\exists(\mathbf{man})(\lambda z.[^2\lambda x.[^3\mathcal{P}_2(\lambda y.P'(x)(y))]^3(z)]^2)]^1$
$\lambda\mathcal{P}_2.\exists(\mathbf{man})(\lambda z.\mathcal{P}_2(\lambda y.[P'(z)(y)]))$

The last proof term left to derive, BOOK, has the following lexical *Map*:

Figure 4



Note that only the SG node *obligatorily* leads to the +DET node, as plural nouns may, but need not, be accompanied by a determiner-- e.g. "*books are one of the best ways to learn*" and "*an informative book is one of the best ways to learn*" vs. *"*book is one of the best ways to* *learn*". As demonstrated by the last example, nouns in the singular form do require determiners (excluding mass nouns e.g. *water*, *money*, etc.). +ADJ and +PP can loop recursively, allowing for an arbitrary number of adjectival and prepositional phrase arguments. In this case, the automaton proceeds SG ➝ +DET ➝ -ADJ ➝ -PP ➝ END, resulting in this final structure; $\lambda\varphi.[\varphi\bullet\text{book}]$; $\lambda\mathcal{P}_4.\mathcal{P}_4(\mathbf{book})$; DET $\Rightarrow$ $NP_{NOM,ACC}$. Terms of the type $NP_{NOM,ACC}$ can appear in all *nominative* and *accusative* positions (i.e. finite subject, nonfinite subject, direct object, indirect object, and prepositional argument). In English, the NP_{NOM}/NP_{ACC} distinction is relevant only for a select subset of the pronouns-- all other nouns are of (or reduce to) the type $NP_{NOM,ACC}$. In other languages such as Croatian, the noun case distinction is maintained through a much more morphologically rich system of nominal inflection.

The singular indefinite article, A, is a two-node lexical entry that, similarly to BOB, does not have any nondeterministic variation in its phrasal structure, and is instantiated in only one possible form-- a; $\lambda x\lambda\mathcal{P}.\exists(x)(\lambda y.\mathcal{P}(y))$; DET, with which BOOK will combine to form the following term; a•book; $\lambda\mathcal{P}.\exists(\mathbf{book})(\lambda y.\mathcal{P}(y))$; $NP_{ACC,NOM}$, where the semantic λ-terms will then β-reduce along the following steps:

14. $\lambda\mathcal{P}_1.[\mathcal{P}_1(\mathbf{book})](\lambda x\lambda\mathcal{P}_2.\exists(x)(\lambda y.\mathcal{P}_2(y)))$
$\lambda x.[\lambda\mathcal{P}_2.\exists(x)(\lambda y.\mathcal{P}_2(y))](\mathbf{book})$
$\lambda\mathcal{P}_2.\exists(\mathbf{book})(\lambda y.\mathcal{P}_2(y))$

The following four terms (15a-d) are generated by the *Pre-Syntactic Automaton* and as such are elements of the proof calculus *Terms* set as per the statement laid out in (11):

15. $\lambda\varphi_1\lambda\varphi_2.[\varphi_1\bullet\text{closed}\bullet\varphi_2]$; $\lambda\mathcal{P}_1\lambda\mathcal{P}_2.\mathcal{P}_1(\lambda x.\mathcal{P}_2(\lambda y.[close(arg_1(x) \cap arg_2(y))]))$; $NP_{NOM} \Rightarrow NP_{ACC} \Rightarrow S$
b. $\lambda\varphi_3.[\varphi_3\bullet\text{book}]$; $\lambda\mathcal{P}_3.\mathcal{P}_3(\mathbf{book})$; DET $\Rightarrow NP_{NOM,ACC}$
c. a; $\lambda w\lambda\mathcal{P}_4.\exists(w)(\lambda z.\mathcal{P}_4(z))$; DET
d. bob; $\lambda\mathcal{P}_5.\mathcal{P}_5(bob)$; $NP_{NOM,ACC}$

Given the terms in (165-d), the proof essentially derives itself, but the individual steps are given anyway as a useful reference:

Figure 5

$\lambda\varphi_1\lambda\varphi_2.[\varphi_1\bullet\text{closed}\bullet\varphi_2]$;
$\lambda\mathcal{P}_1\lambda\mathcal{P}_2.\mathcal{P}_1(\lambda x.\mathcal{P}_2(\lambda y.[\mathit{close}(\mathit{arg}_1(x) \cap \mathit{arg}_2(y))]))$;
$NP_{NOM} \Rightarrow NP_{ACC} \Rightarrow S$ bob; $\lambda\mathcal{P}_5.\mathcal{P}_5(\mathit{bob})$; $NP_{NOM,ACC}$

| |
| | $\lambda\varphi_3.[\varphi_3\bullet\text{book}]$; a; $\lambda w\lambda\mathcal{P}_4.\exists(w)(\lambda z.\mathcal{P}_4(z))$;
| | $\lambda\mathcal{P}_3.\mathcal{P}_3(\textbf{book})$; DET $\Rightarrow NP_{NOM,ACC}$ DET
| | --

$\lambda\varphi_2.[\text{bob}\bullet\text{closed}\bullet\varphi_2]$; | |
$\lambda\mathcal{P}_2.\mathcal{P}_2(\lambda y.[\mathit{close}(\mathit{arg}_1(\mathit{bob}) \cap \mathit{arg}_2(y))])$; a•book; $\lambda\mathcal{P}_4.\exists(\textbf{book})(\lambda z.\mathcal{P}_4(z))$;
$NP_{ACC} \Rightarrow S$ $NP_{NOM,ACC}$

| |
bob•closed•a•book; $\exists(\textbf{book})(\lambda z.[\mathit{close}(\mathit{arg}_1(\mathit{bob}) \cap \mathit{arg}_2(z))])$; S

As one can see, though the *Pre-Syntactic Automaton* may be a complex system, it provides several clear advantages. Primarily, it accounts for the nagging coincidence that nearly all other systems of formal syntactic analysis must accept as a linguistic fact; multiple lexical entries must be posited for homophonous and semantically (near-)synonymous forms of the same lexemes, in order to account for the various possible phrasal structures and morphosyntactic phenomena associated with a given lexeme, e.g. the differing valences of the English CLOSE in its active, middle, and passive manifestations. While a given lexeme may enumerate a plethora of terms, almost every other lexeme does the same, and as such it is not an unusual exception to the general pattern. Additionally, by forcing one to consider the impact each minute variation in the phonological, semantic, or tectogrammatical sectors for a given phrasal structure may have on any of the other two elements, a syntactician working within the NECG framework is severely restricted in their ability to "hand-wave" and ignore or abstract away from small details, but is also presented with mechanisms both powerful and fine-tuned enough that such "hand-waving" is not necessary. Finally, as exemplified in figure 5, while pre-syntactic derivations may be intricate, these lead to exceedingly self-explanatory and straightforward derivational proofs.

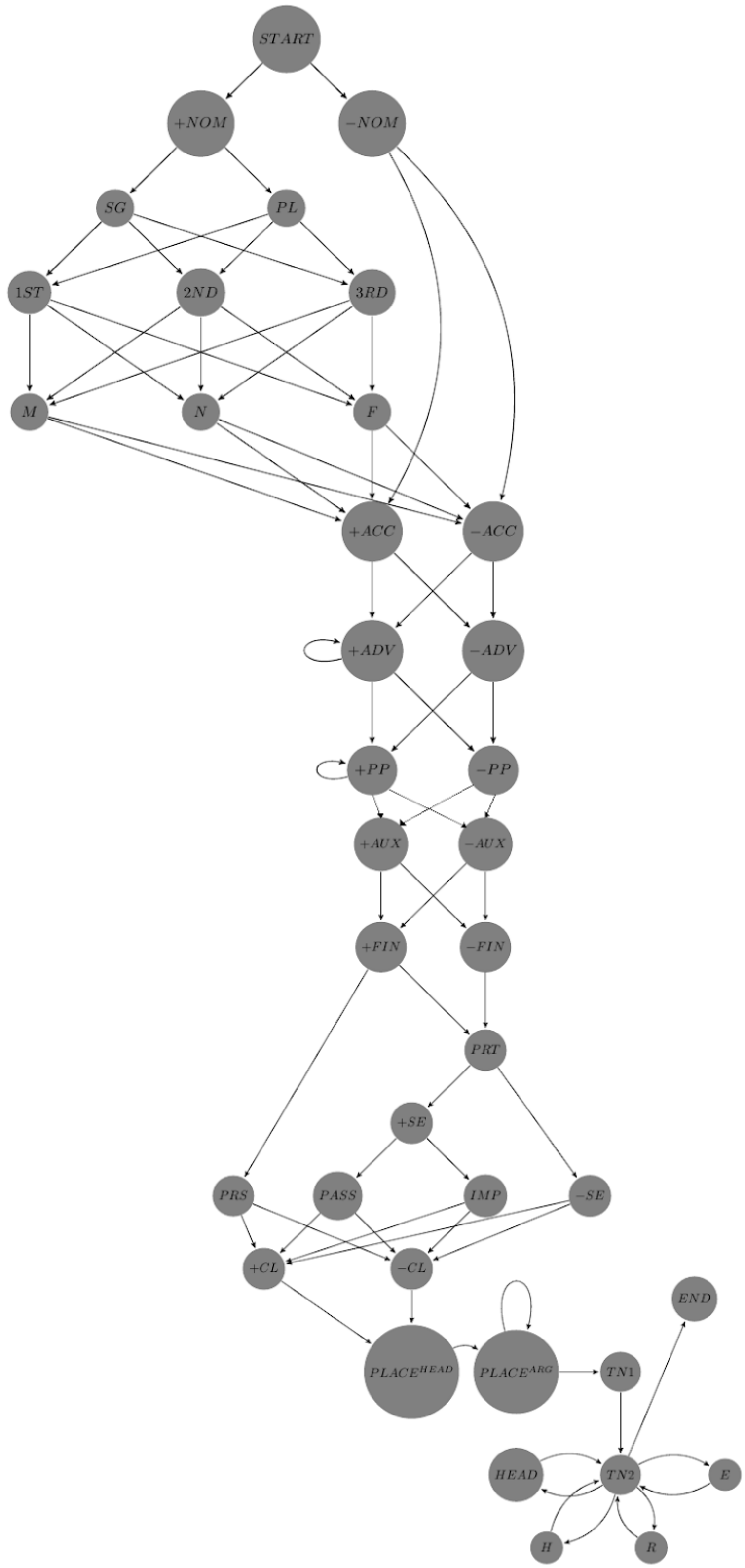
(2.1.2) Croatian Active Voice

In order to derive the Croatian active voice construction first presented in example(1), "*Marko je zatvorio knjigu*" ("*Marko closed the book*"), four proof terms that first must be obtained via the *Pre-Syntactic Automaton*, corresponding to the isolated phonological strings in (16a-d):

16a. *Marko* - (“*Marko*”)
b. *je* - (auxiliary “*to be*”)
c. *zatvorio* - (masculine singular participial form of “*close*”)
d. *knjigu* - (accusative singular form of “*book*”)

As in the corresponding English derivation, *zatvorio* will be derived first. Its directed graph *Map* is given in figure 6 on the following page:

Figure 6


START
+NOM
−NOM
SG
PL
1ST
2ND
3RD
M
N
F
+ACC
−ACC
+ADV
−ADV
+PP
−PP
+AUX
−AUX
+FIN
−FIN
PRT
+SE
PRS
PASS
IMP
−SE
+CL
−CL
END
PLACE^HEAD
PLACE^ARG
TN1
HEAD
TN2
E
H
R

Like all other lexical entries/algorithms, ZATVORITI begins at its START node, which, unlike the English lexeme CLOSE, does not add positional tuples to the set of *global variables*, instead, there is merely an unordered set of arguments, *args* (all elements of *args* are triples of the form $<\varphi, \Sigma, X>$). While English has a fairly rigid word order, and the semantic role of each argument in a given term is determined via its position in the linear order, Croatian word order is much more fluid. This necessitates an unordered argument set that, as a consequence of the system's nondeterminism, will result in the enumeration of all possible linear sequences of the *args* set's elements. As previously mentioned in section (1), Croatian clitics have a fixed position; they must be in the second phrasal position. As such, a variable *clitic*-- a triple of the form $<\varphi, \Sigma, X>$ -- serves to reserve this crucial second position. Finally, the same ARG_P variable that was first introduced in CLOSE is also added to the set of *global variables* by the *instruction function* of ZATVORITI.

Rather than selecting for subject, or lack thereof, as in CLOSE, the system's first move in ZATVORITI's algorithm determines whether or not it will have a nominative argument (+NOM and -NOM). In this case, the automaton selects +NOM as its first move, and proceeds to select morphosyntactic agreement features for its nominative argument in a similar fashion to CLOSE, with the exception that ZATVORITI additionally selects gender agreement features. In this case, the path followed is +NOM → SG → 3RD → M, agreeing with, and selecting for, the masculine third person singular NP_{NOM} *Marko*. The three gender agreement nodes add the nominative NP argument, as they are the provide the final morphosyntactic information necessary to select for a correctly agreeing NP subtype. In this instance, $<\varphi_i, x, NP_{NOM,SG,3,M}>$ is added as an element of the set *args*.

The system may now proceed to either +ACC or -ACC, determining whether or not the output term will select for an accusative NP argument. In this case, the system does move from M to +ACC, adding an element $<\varphi_j, y, NP_{ACC}>$ to the *args* set. From this point, it proceeds along the path -ADV ⟶ -PP ⟶ +AUX. The *instruction function* of +AUX sets the variable *clitic* equal to $<\varphi_i, \mathcal{P}_j, Aux_{3, SG}>$, which is then changed to $<\varphi_i, \mathcal{P}_j, Aux_{3,SG,FIN}>$, after the automaton makes its next move to +FIN, which additionally sets the *terminal reducing type* X_T to S_{FIN}, and sets ARG_P = $arg_1(x) \cap arg_2(y)$, where x is the semantic variable representing the NP_{NOM} argument, and y represents the NP_{ACC} argument. The system's next move proceeds to PRT (participle), which sets φ_{THIS} equal to $\beta_{ZATVORITI}(P^{TCP}\bullet M\bullet S^{G})$-- "*zatvorio*". From PRT, the next move in this derivation is to -SE, which essentially just serves to prevent the automaton from proceeding to the PASS or IMP nodes corresponding to the set of constructions that utilize the third person reflexive/impersonal pronoun SE.

-CL and +CL are restricted by their *condition functions* to only be available as a possible next move to the system when *clitic* = ø and *clitic* ≠ ø, respectively. Essentially, if the *clitic* variable is valued, the system must proceed to +CL, and if not, it must proceed to -CL. In this case, *clitic* = $<\varphi_i, \mathcal{P}_j, Aux_{3,SG,FIN}>$, and so the system obligatorily proceeds to +CL, whose *instruction function* mandates that u'_2 = *clitic*, and sets *occupied* = {2} obeying a general rule stating that clitics almost always appear in the second position of a phrase in Croatian (Schütze 1994). *occupied* $\subseteq \mathbf{N}$ denotes every position in u' that is filled by either an argument or φ_{THIS}.

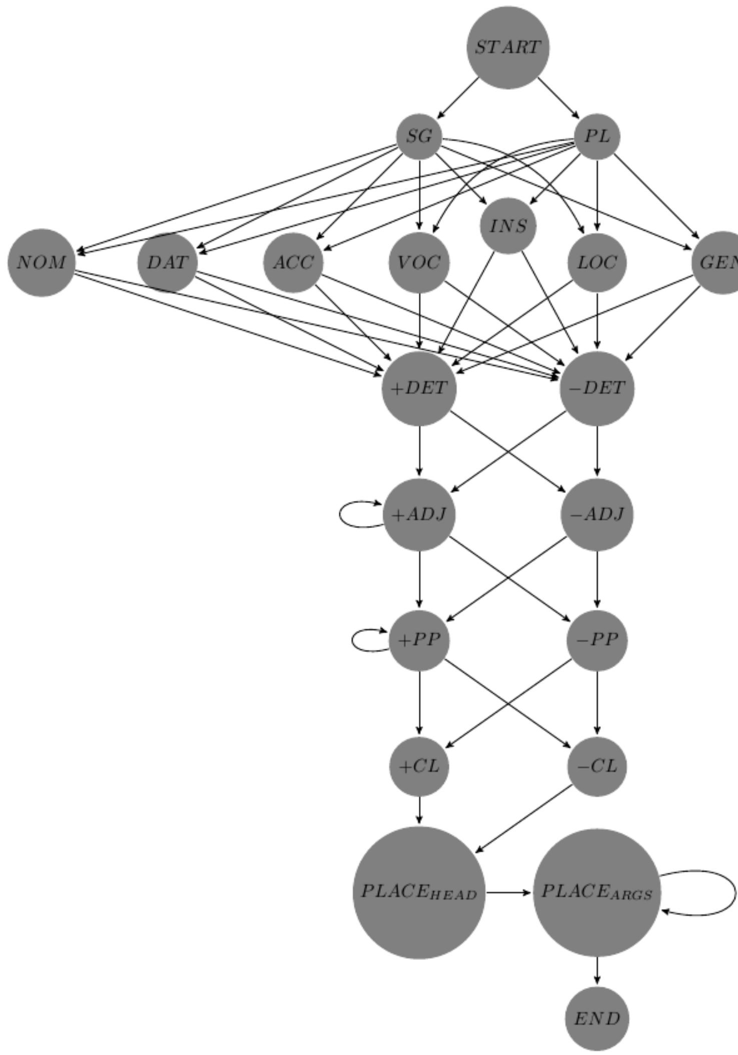
The system now proceeds to $\text{PLACE}^{\text{HEAD}}$, which randomly selects an element of the subset of the natural numbers [1, n] \ *occupied*, where n = |args| + |*clitic*| + 1. In this case, n = 4, representing the sum of the number of freely placeable arguments-- 2, the number of clitic arguments-- 1, and adds an additional 1, reserving a position for φ_{THIS}. This set contains the indices of all unfilled positions in the tuple u' (where position 2 begins already occupied by the auxiliary clitic). In this case, $\text{PLACE}^{\text{HEAD}}$ selects the number 3 (*Marko je* *zatvorio* *knjigu*), and adds it to the *occupied* set, which then becomes the two-member set {2, 3}.

At this point, *args* = $\{\langle\varphi_k, x, NP_{3,SG,M,NOM}\rangle, \langle\varphi_n, y, NP_{ACC}\rangle\}$, and the system proceeds to the $\text{PLACE}^{\text{ARGS}}$ node, which loops recursively, extending the process laid out in $\text{PLACE}^{\text{HEAD}}$ to each element *a* of the *args* set, each time setting *args* = *args* \ {*a*}. The system may proceed to TN1 only once *args* = ø. Once the automaton has moves to TN1 in the current derivation, $u' = \langle\langle\varphi_1, x, NP_{3,SG,M,NOM}\rangle, \langle\varphi_2, \mathcal{P}_2, Aux_{3,SG,FIN}\rangle$, "*zatvorio*", $\langle\varphi_3, x, NP_{ACC}\rangle\rangle$. From the TN1 node onward, the derivation exactly mirrors that of CLOSE in the English active construction derived in section (2.1.1), eventually yielding the following term (any of the five other possible orderings of the elements φ_1, φ_3, and "*zatvorio*" are also derivable);

19. $\lambda\varphi_1\lambda\varphi_2\lambda\varphi_3.[\varphi_1\bullet\varphi_2\bullet\text{zatvorio}\bullet\varphi_3]$; $\lambda\mathcal{P}_1\lambda\mathcal{P}_2\lambda\mathcal{P}_3.\mathcal{P}_2(\mathcal{P}_1(\lambda x.\mathcal{P}_3(\lambda y.[\mathit{close}(\mathit{arg}_1(x) \cap \mathit{arg}_2(y))])))$; $NP_{NOM,3,M,SG} \Rightarrow Aux_{3,SG,FIN} \Rightarrow NP_{ACC} \Rightarrow S$

Next, the term knjigu; $\lambda\mathcal{P}.\exists(\mathbf{book})(\lambda z.\mathcal{P}(z))$; NP_{ACC} is derived from the lexeme KNJIG, represented by the directed graph in figure 7 below:

Figure 7



From the START node, the automaton may proceed to either SG or PL, corresponding to singular and plural noun inflection, respectively. *knjigu* is singular, and as such the automaton moves to SG in this derivation. The NOM, DAT, ACC, VOC, LOC, INS, and GEN nodes correspond to the seven Croatian noun cases, and in this case, the next move is to ACC (accusative), deriving a feminine (KNJIGA is inherently feminine) noun with singular accusative inflection, and setting $X_T = NP_{ACC}$, and $\varphi_{THIS} = \beta_{KNJIGA}(S^G \bullet A^{CC})$-- "*knjigu*". From any of the seven nominal case nodes, the system may proceed either +DET or -DET, owing to the fact that, as in many Slavic languages, determiners are not obligatory in Croatian (Progovac 1982). The now familiar +ADJ, -ADJ, +PP, and -PP nodes function just as in the English BOOK, with the exception that +PP and +ADJ add prepositional and adjectival arguments, respectively, to the unordered set *args*, as in ZATVORITI.

Regardless, in this case the system proceeds -DET ⟶ -ADJ ⟶ -PP, where -ADJ and -PP do not make any changes to the set of *global variables*. -DET existentially quantifies KNJIGA, changing its semantics to $\lambda\mathcal{P}.\exists(\mathbf{book})(\lambda z.\mathcal{P}(z))$. The system then proceeds to -CL, as there are no clitic arguments in the intended structure (21a), and as such *clitic* = ø. $\text{PLACE}^{\text{HEAD}}$ only has one position available to in which to place φ_{THIS}, as this instantiation of KNJIGA does not take any arguments, and $\text{PLACE}^{\text{ARGS}}$ makes no change to the set of *global variables*. By the time the system reaches the END node, u' = <"*knjigu*">, $\Sigma_T = \lambda\mathcal{P}.\exists(\mathbf{book})(\lambda z.\mathcal{P}(z))$, and $X_T = NP_{ACC}$. Via stages 2-8 of the *Pre-Syntactic Automaton*, the final term is obtained:

20a. knjigu; $\lambda\mathcal{P}.\exists(\mathbf{book})(\lambda z.\mathcal{P}(z))$; NP_{ACC}

The *Map* of the lexeme MARKO is extremely similar to that of the English BOB, with the exception that it allows for noun case inflection. In any case, it is fairly trivial to obtain the following term, and so for the sake of brevity, the visual diagram of the directed graph is not given.

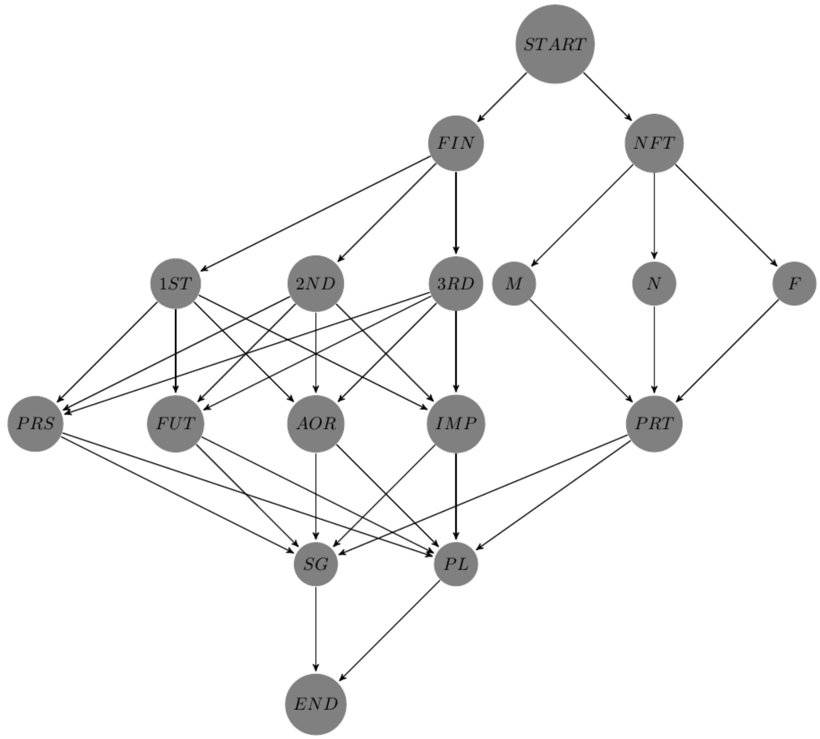
20b. marko; $\lambda\mathcal{P}.\mathcal{P}(marko)$; $NP_{NOM,3,SG,M}$

The final term necessary to derive the Croatian active construction in question is the auxiliary clitic *je*:

20c. je; $\lambda\mathcal{P}.\mathcal{P}$; $Aux_{FIN,3,SG}$

Note that semantically, this term is essentially just an n^{th}-Order identity function. In other words, it makes no changes to the underlying semantic interpretation of any given term that it takes as an argument. *je* is the finite third person singular form of the lexeme BITI, and can be derived from its *Map*, given in figure 8 below:

Figure 8



In this instance, the system proceeds from the START node to FIN, setting the boolean *finite* = 1 (*finite* is a *global variable*). As *je* is the third person singular form, the system now proceeds to 3RD, setting its person agreement value to 3. The verb BITI has four finite forms, aorist (AOR), present (PRS), future II (FUT) and imperfect (IMP). *je* is inflected for the present tense, and as such the automaton moves from 3RD to PRS, then to SG, which sets $\varphi_{THIS} = \beta_{BITI}(P^{RS}\bullet 3\bullet S^{G})$ (= "*je*"). At this point, the system only has one possible available move, and proceeds to the END node, terminating this phase of the derivation, and "setting the dominoes", in a sense, to eventually yield the final output term:

20d. je; $\lambda\mathcal{P}.\mathcal{P}$; $Aux_{FIN,3,SG}$

With these four proof terms derived from the *Pre-Syntactic Automaton* and given in (20a-d), the actual derivational proof itself is again an extremely straightforward process:

Figure 9

$\lambda\varphi_1\lambda\varphi_2\lambda\varphi_3.[\varphi_1\bullet\varphi_2\bullet\text{zatvorio}\bullet\varphi_3]$;
$\lambda\mathcal{P}_1\lambda\mathcal{P}_2\lambda\mathcal{P}_3.\mathcal{P}_2(\mathcal{P}_1(\lambda x.\mathcal{P}_3(\lambda y.[close(arg_1(x) \cap arg_2(y))])))$; marko; $\lambda\mathcal{P}_4.\mathcal{P}_4(marko)$;
$NP_{NOM,3,SG,M} \Rightarrow Aux_{FIN,3,SG} \Rightarrow NP_{ACC} \Rightarrow S$ $NP_{NOM,3,SG,M}$

||
$\lambda\varphi_2\lambda\varphi_3.[\text{marko}\bullet\varphi_2\bullet\text{zatvorio}\bullet\varphi_3]$;
$\lambda\mathcal{P}_2\lambda\mathcal{P}_3.\mathcal{P}_2(\mathcal{P}_3(\lambda y.[close(arg_1(marko) \cap arg_2(y))]))$; je; $\lambda\mathcal{P}_5.\mathcal{P}_5$;
$Aux_{FIN,3,SG} \Rightarrow NP_{ACC} \Rightarrow S$ $Aux_{FIN,3,SG}$

||
||
$\lambda\varphi_3.[\text{marko}\bullet\text{je}\bullet\text{zatvorio}\bullet\varphi_3]$;
$\lambda\mathcal{P}_3.\mathcal{P}_3(\lambda y.[close(arg_1(marko) \cap arg_2(y))])$;
$NP_{ACC} \Rightarrow S$ knjigu; $\lambda\mathcal{P}_5.\exists(book)(\lambda z.\mathcal{P}_5(z))$; NP_{ACC}

||
marko•je•zatvorio•knjigu; $\exists(book)(\lambda z.[close(arg_1(marko) \cap arg_2(z))])$; S

Any other ordering of the three strings *marko*, *zatvorio*, and *knjigu* can be obtained by virtue of the nondeterminism of the system (as a clitic, *je* must remain in second position), resulting in a different implicational and λ-term ordering, but still arriving at the same semantic interpretation (truth-conditionally, at least-- different orderings may result in different topicalization/focus effects, which could be an interesting route for further research).

(2.2) Passive Voice

In both the English ("*a book was closed* [*by Bob*]") and Croatian ("*zatvorila se knjiga* [*od Marka*]") passive constructions, the subject/nominative argument is assigned the semantic role canonically given to the object/accusative argument, the object/accusative argument does not appear in the structure, and the canonical semantic subject argument can be reintroduced via the prepositions *by* and *od* in English and Croatian, respectively. As the following two sections will demonstrate, these constructions are fairly easily accounted for in the *Nondeterministically Enumerated Categorial Grammar* framework.

(2.2.1) English Passive Voice

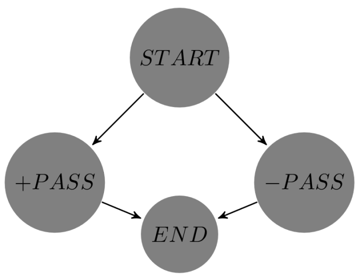
Beginning with English, two of the terms derived in section (2.1.1) can be recycled, namely; *a book* and *Bob*. The passive auxiliary *was*, the preposition *by*, and the passive form of CLOSE must be derived pre-syntactically in order to construct the final derivational proof.

The auxiliary BE is tectogrammatically similar to the Croatian BITI, in that it does not select for any arguments, and merely inflects for tense, person, and number values. For the sake of brevity, the term for *was* is provided here:

21a. was; $\lambda\mathcal{P}.\mathcal{P}$; $\mathrm{Aux}_{\mathrm{FIN,3,SG}}$

Figure 10 below is the digraph *Map* for the lexeme BY:

Figure 10

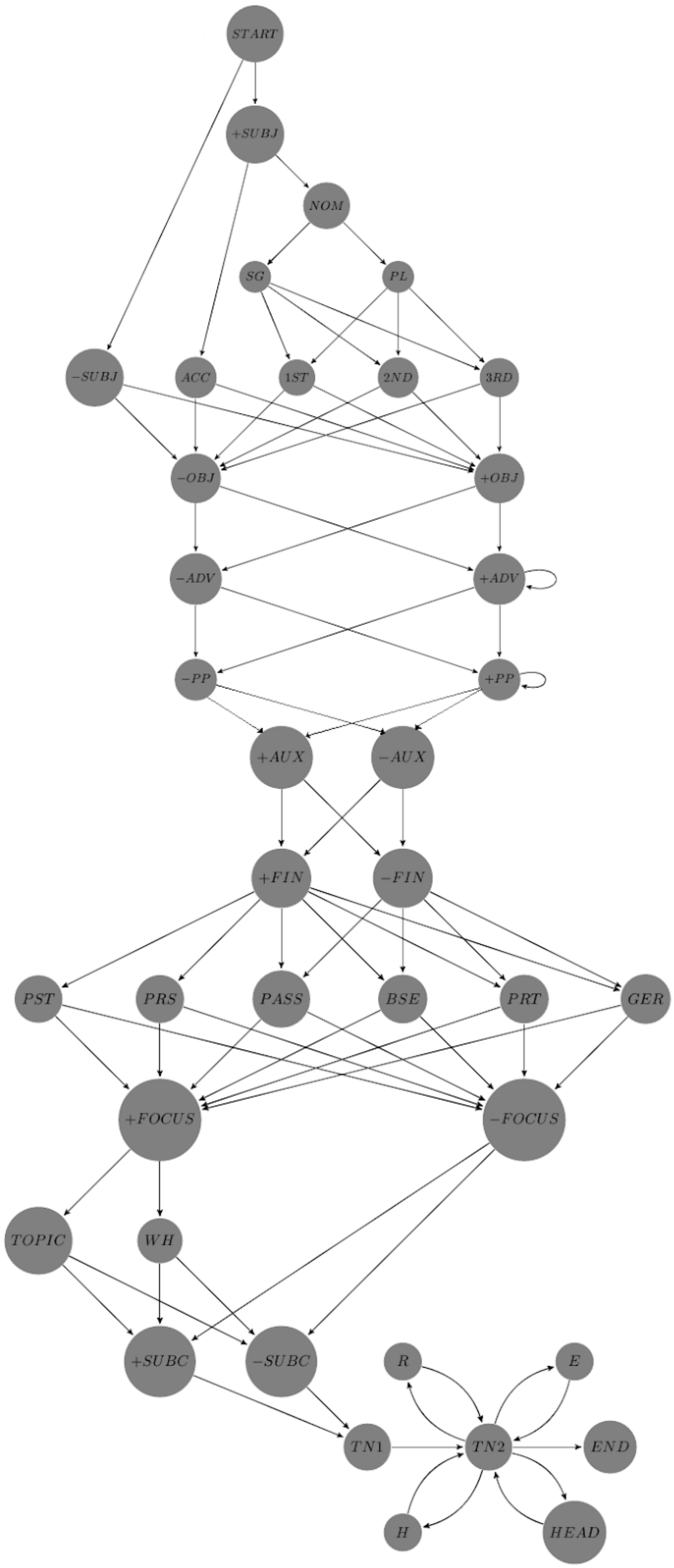


From the START node, the automaton may proceed to either the +PASS or -PASS nodes, where the +PASS node yields the following semantic interpretation; $\lambda\mathcal{P}_1\lambda\mathcal{P}_2.\mathcal{P}_1(\lambda x.\mathcal{P}_2(arg_1(x)))$, and -PASS results in $\lambda\mathcal{P}_1\lambda\mathcal{P}_2.\mathcal{P}_1(\lambda x.\mathcal{P}_2(by(x)))$, where the functional predicate *by*(x) confers some kind of abstract locational semantic interpretation. In this case, however, the system moves to +PASS, which additionally changes the *terminal reducing type* variable X_T to $PP_{\text{BY-PASS}}$ (i.e. the passive form of *by*) and subsequently allowing this preposition to reintroduce the *primary argument*, which is not permitted in its canonical subject position by the passive construction. The only possible move from +PASS is to END, and the obligatory *Pre-Syntactic Automaton* stages 2-8 result in the following proof term:

21b. $\lambda\varphi.[\mathrm{by}\bullet\varphi]$; $\lambda\mathcal{P}_1\lambda\mathcal{P}_2.\mathcal{P}_1(\lambda x.\mathcal{P}_2(arg_1(x)))$; $NP_{\mathrm{ACC}} \Rightarrow PP_{\text{BY-PASS}}$

There now remains one final term to be derived in order to begin the proof for the English passive construction; the passive form of CLOSE. The *Map* of CLOSE, originally given in figure 2, is again given in figure 11 below for reference:

Figure 11


START
+SUBJ
NOM
SG
PL
−SUBJ
ACC
1ST
2ND
3RD
−OBJ
+OBJ
−ADV
+ADV
−PP
+PP
+AUX
−AUX
+FIN
−FIN
PST
PRS
PASS
BSE
PRT
GER
+FOCUS
−FOCUS
TOPIC
WH
+SUBC
−SUBC
R
E
TN1
TN2
END
H
HEAD

This derivation begins much like that of the active form, starting along the following path; +SUBJ → NOM → SG → 3RD, but the two derivations diverge when, only in the current derivation, the system proceeds to -OBJ. Recall that at this point, the subject variable, x, is still stored in *arg_subj* (a *global variable*), and has not yet been assigned one of the two available semantic roles, arg_1 or arg_2. The system then proceeds to -ADV, and next to +PP, adding a prepositional-phrase argument to the *global variable* tuple *pp*. From here, the system proceeds to +AUX, adding an auxiliary-type argument to the *global variable* tuple *aux*, and introducing the boolean variable *auxiliary* = 1 to the set of *global variables*. The next move advances to +FIN, introducing the boolean variable *finite* = 1 to the set of *global variables*.

The node PASS (passive) is restricted by its *condition function*, and is only a valid move if *auxiliary* = 1 and *length*(*obj*) = 0 (i.e. there is no object). These *conditions* are met, and PASS is selected, which sets the argument of the predicate CLOSE such that ARG_P = ARG_P ∩ arg_2(*arg_subj*), sets the variable X_T (the construction's *terminal reducing type*) to S_{PASS}, and $\varphi_{THIS} = \beta_{CLOSE}(P^{TCP})$-- "*closed*". Additionally, the PASS node sets the tectogrammatical type of the auxiliary (the only element of the *aux* tuple) to Aux_{BE}. At this point in the derivation, ARG_P = R ∩ arg_2(x), where x is the semantic entity-type variable of the subject argument, and R is the relational predicate-type variable represented by the PP argument (stored in u' as $<\varphi_3$, R, PP>). Proceeding to -FOCUS, -SUBC, TN1, etc., the system eventually reaches the END node, at which point u' = $<<\varphi_1, x, NP_{NOM,3,SG}>, <\varphi_2, \mathcal{P}_2, NP_{NOM,3,SG}>$, "*closed*", $<\varphi_3$, R, PP>>, Σ_T = *close*(R ∩ arg_2(x)), and $X_T = S_{PASS}$. Via the subsequent stages (i.e. 2-8) of the *Pre-Syntactic Automaton*, this final term is obtained:

21c. $\lambda\varphi_1\lambda\varphi_2\lambda\varphi_3.[\varphi_1 \bullet \varphi_2 \bullet \text{closed} \bullet \varphi_3]; \lambda\mathcal{P}_1\lambda\mathcal{P}_2\lambda\mathcal{P}_3.\mathcal{P}_2(\mathcal{P}_1(\lambda x.\mathcal{P}_3(\lambda R.[\mathit{close}(\mathit{arg}_2(x) \cap R)])))$;
$NP_{3,SG,NOM} \Rightarrow Aux_{BE,FIN,3,SG} \Rightarrow PP \Rightarrow S$

With a combination of the terms derived in section (2.1.1), and those given in (21a-c), the derivational proof of the English passive construction "*a book was closed by Bob*" will proceed as follows in figure 12:

Figure 12

$\lambda\varphi_1\lambda\varphi_2\lambda\varphi_3.[\varphi_1\bullet\varphi_2\bullet\text{closed}\bullet\varphi_3]$;
$\lambda\mathcal{P}_1\lambda\mathcal{P}_2\lambda\mathcal{P}_3.\mathcal{P}_2(\mathcal{P}_1(\lambda x.\mathcal{P}_3(\lambda R.[\mathit{close}(\mathit{arg}_2(x) \cap R)])))$;
$NP_{3,SG,NOM} \Rightarrow Aux_{BE,FIN,3,SG} \Rightarrow PP \Rightarrow S$

a•book; $\lambda\mathcal{P}_4.\exists(\mathbf{book})(\lambda y.\mathcal{P}_4(y))$;
$NP_{3,SG,NOM,ACC}$

$\lambda\varphi_2\lambda\varphi_3.[\text{a}\bullet\text{book}\bullet\varphi_2\bullet\text{closed}\bullet\varphi_3]$;
$\lambda\mathcal{P}_2\lambda\mathcal{P}_3.\mathcal{P}_2(\exists(\mathbf{book})(\lambda y.\mathcal{P}_3(\lambda R.[\mathit{close}(\mathit{arg}_2(y) \cap R)])))$;
$Aux_{BE,FIN,3,SG} \Rightarrow PP \Rightarrow S$

was; $\lambda\mathcal{P}_5.\mathcal{P}_5$;
$Aux_{BE,FIN,3,SG}$

$\lambda\varphi_4.[\text{by}\bullet\varphi_4]$;
$\lambda\mathcal{P}_6\lambda\mathcal{P}_7.\mathcal{P}_6(\lambda z.\mathcal{P}_7(\mathit{arg}_1(z)))$;
$NP_{ACC} \Rightarrow PP_{BY\text{-}PASS}$

bob; $\lambda\mathcal{P}_8.\mathcal{P}_8(\mathit{bob})$;
$NP_{NOM,ACC}$

----

by•bob; $\lambda\mathcal{P}_7.\mathcal{P}_7(\mathit{arg}_1(\mathit{bob}))$; $PP_{BY\text{-}PASS}$

$\lambda\varphi_3.[\text{a}\bullet\text{book}\bullet\text{was}\bullet\text{closed}\bullet\varphi_3]$;
$\lambda\mathcal{P}_3.\exists(\mathbf{book})(\lambda y.\mathcal{P}_3(\lambda R.[\mathit{close}(\mathit{arg}_2(y) \cap R)]))$;
$PP \Rightarrow S$

----

a•book•was•closed•by•bob; $\exists(\mathbf{book})(\lambda y.[\mathit{close}(\mathit{arg}_2(y) \cap \mathit{arg}_1(\mathit{bob}))])$; S

In order to obtain the passive construction *without* the arg_1 being re-introduced by a *by*-prepositional phrase (i.e. "*a book was closed*"), the system simply moves from -OBJ to -PP, as opposed to +PP, which (all else being the same) results in the following proof calculus term:

22. $\lambda\varphi_1\lambda\varphi_2.[\varphi_1\bullet\varphi_2\bullet\text{closed}]$; $\lambda\mathcal{P}_1\lambda\mathcal{P}_2.\mathcal{P}_2(\mathcal{P}_1(\lambda x.[\mathit{close}(\mathit{arg}_2(x))]))$;
$NP_{3,SG,NOM} \Rightarrow Aux_{BE,FIN,3,SG} \Rightarrow S$

(2.2.2) Croatian Passive Voice

As with the English passive derivation, some of the terms necessary to derive the Croatian passive construction "*Zatvorila se knjiga od Marka*" have already been (almost) derived in section (2.1.2). To obtain *knjiga*, the system simply moves to the NOM node instead of ACC, which ultimately results in the following proof term:

23a. knjiga; $\lambda\mathcal{P}.\exists(\mathbf{book})(\lambda x.\mathcal{P}(x))$; $NP_{NOM,3,SG,F}$

Similarly, *Marka* (the genitive form of MARKO) is obtained when the automaton selects the GEN (genitive)-- as opposed to NOM-- node, and the following term is derived:

23b. marka; $\lambda\mathcal{P}.\mathcal{P}(marko)$; $NP_{GEN,3,SG,M}$

The preposition OD contains the same four nodes as its synonymous English counterpart, *by*; namely, START, +PASS, -PASS, and END-- and +PASS and -PASS confer the same semantics as their respective BY counterparts. So, in order to obtain a term of the type $PP_{OD\text{-}PASS}$, the automaton moves along this path; START → +PASS → END, which will derive the following term:

23c. $\lambda\varphi.[od\bullet\varphi]$; $\lambda\mathcal{P}_1\lambda\mathcal{P}_2.\mathcal{P}_1(\lambda x.\mathcal{P}_2(arg_1(x)))$; $NP_{GEN} \Rightarrow PP_{OD\text{-}PASS}$

The clitic SE is hypothesised to be a two-node lexeme that only has one possible phrasal structure, given in (23d) below:

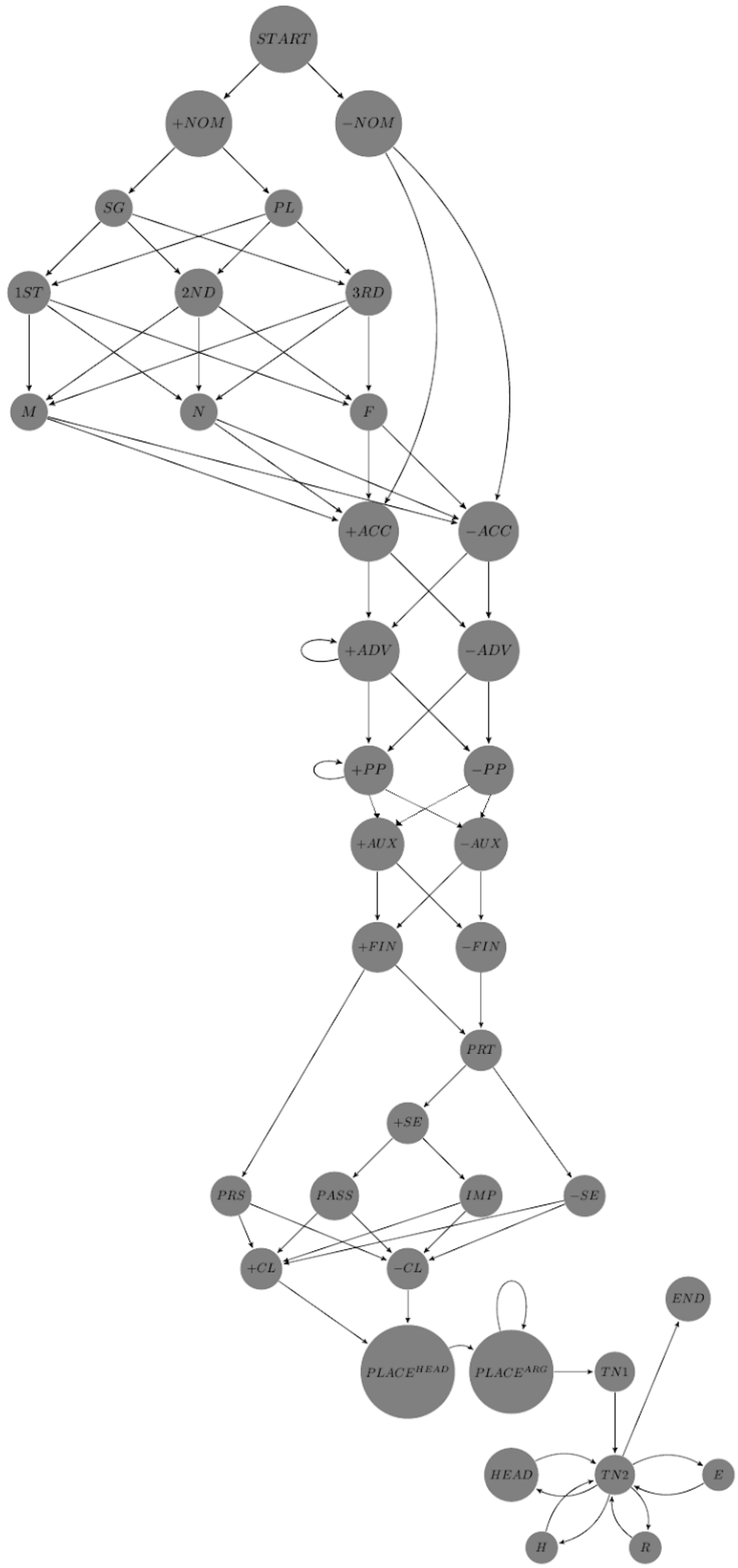
23d. se; $\lambda\mathcal{P}.\mathcal{P}$; SE

The reasoning behind this supposition is somewhat simple; there exist a large amount of constructions across the BCS continuum that utilize the SE clitic (Rivero 2001), which seems to have varying interpretations in each construction, which are somewhat related in that they all seem to affect the agent-like argument. This effect, however, can range from valence-reduction (impersonal and passive) and backgrounding (when reintroduced via a passive *od*-phrase), to seemingly functioning as a reflexive or reciprocal pronoun.

The central hypothesis put forward in this analysis is that SE has no real semantic interpretation in isolation-- if one is an adherent to the view that language is purely a system which enumerates a set of form-meaning pairs, it does not seem implausible that SE is simply a prosodic *non-canonical marker*. In other words, it marks a given construction as having a different mapping between syntactic arguments and semantic roles than in its canonical form. The specific pattern of alterations to this mapping is then independently determined by each respective class of lexical entries and morphosyntactic constructions that are able to select for SE (hence the semantics of SE being an n^{th}-Order *identity function*).

With the terms given in (23a-d), the only remaining piece of the puzzle is the actual passive form of ZATVORITI. The *Map* of ZATVORITI, which originally appeared in figure 6, is again given in figure 13 for reference:

Figure 13


START
+NOM
−NOM
SG
PL
1ST
2ND
3RD
M
N
F
+ACC
−ACC
+ADV
−ADV
+PP
−PP
+AUX
−AUX
+FIN
−FIN
PRT
+SE
PRS
PASS
IMP
−SE
+CL
−CL
END
PLACE^HEAD
PLACE^ARG
TN1
HEAD
TN2
E
H
R

Again, analogous to the passive derivation of the English verb CLOSE, this pre-syntactic derivation of the passive form of ZATVORITI begins in a very similar fashion to its active counterpart. The automaton begins with the following path; START → +NOM → SG → 3RD → F, where the only difference up to this point is that it selects for feminine gender agreement (to agree with *knjiga*), as opposed to the masculine selected by the active derivation in (23). Here, the two derivations truly begin to diverge. In this case, the next move is to -ACC (as the passive construction does not permit an accusative argument), then -ADV, and finally to +PP (in order to allow the possibility of introducing the $PP_{OD\text{-}PASS}/arg_1$ argument). As in the active derivation, +AUX → +FIN → PRT are the next three moves made by the system, and the *clitic* variable is set to a finite auxiliary-type (i.e. *clitic* = $<\varphi_i, \mathcal{P}_j, Aux_{FIN}>$). However, following PRT, this derivation the proceeds to the +SE node, which paves the way for either a passive or impersonal construction. In addition, the *instruction function* of +SE sets *clitic* = $<\varphi_i, \mathcal{P}_j, SE>$, effectively overwriting the auxiliary clitic, and accounting for its absence in both the passive and impersonal. After +SE, the automaton continues to PASS (passive), whose *condition function* only returns true if there exists a nominative NP, and there does not exist an accusative NP, in the *args* set (*global variable*); a requirement which is satisfied by the present state of the system. PASS sets $\varphi_{THIS} = \beta_{ZATVORITI}(P^{TCP} \bullet F \bullet S^{G})$ and ARG_P = ARG_P ∩ arg_2(x), where, similarly to the English passive construction, x is the semantic entity-type variable of the nominative argument, and R is the relational predicate-type variable represented by the PP argument. As in the English passive derivation, the passivization process sets $X_T = S_{PASS}$. From here, the derivation follows the same steps as in the active derivation of *zatvorio*, and by the END node, u' = $<$"*zatvorila*", $<\varphi_1, \mathcal{P}_1, SE>, <\varphi_2, \mathcal{P}_2, NP_{NOM,3,SG,F}>, <\varphi_3, \mathcal{P}_3, PP>>$ (where the altered word order again falls out from the nondeterminism of the system), $\Sigma_T = close(R \cap arg_2(x))$, and the *terminal reducing type* $X_T = S_{PASS}$. The final term outputted by the *Pre-Syntactic Automaton* is then the following:

23e. $\lambda\varphi_1\lambda\varphi_2\lambda\varphi_3.[\text{zatvorila} \bullet \varphi_1 \bullet \varphi_2 \bullet \varphi_3]; \lambda\mathcal{P}_1\lambda\mathcal{P}_2\lambda\mathcal{P}_3.\mathcal{P}_1(\mathcal{P}_2(\lambda x.\mathcal{P}_3(\lambda R.[close(arg_2(x) \cap R)])));$
$SE \Rightarrow NP_{NOM,3,SG,F} \Rightarrow PP \Rightarrow S$

Given the terms in (23a-e), the proof in figure 14 can be derived:

Figure 14

$\lambda\varphi_1\lambda\varphi_2\lambda\varphi_3.[\text{zatvorila}\bullet\varphi_1\bullet\varphi_2\bullet\varphi_3];$
$\lambda\mathcal{P}_1\lambda\mathcal{P}_2\lambda\mathcal{P}_3.\mathcal{P}_1(\mathcal{P}_2(\lambda x.\mathcal{P}_3(\lambda R.[\mathit{close}(\mathit{arg}_2(x) \cap R)])));$
$\text{SE} \Rightarrow \text{NP}_{\text{NOM,3,SG,F}} \Rightarrow \text{PP} \Rightarrow \text{S}$ se; $\lambda\mathcal{P}_4.\mathcal{P}_4$; SE

||
||
||
||

$\lambda\varphi_2\lambda\varphi_3.[\text{zatvorila}\bullet\text{se}\bullet\varphi_2\bullet\varphi_3];$
$\lambda\mathcal{P}_2\lambda\mathcal{P}_3.\mathcal{P}_2(\lambda x.\mathcal{P}_3(\lambda R.[\mathit{close}(\mathit{arg}_2(x) \cap R)]));$ knjiga; $\lambda\mathcal{P}_5.\exists(\mathbf{book})(\lambda y.\mathcal{P}_5(y));$
$\text{NP}_{\text{NOM,3,SG,F}} \Rightarrow \text{PP} \Rightarrow \text{S}$ $\text{NP}_{\text{NOM,3,SG,F}}$

|| $\lambda\varphi_4.[\text{od}\bullet\varphi_4];$ marka;
|| $\lambda\mathcal{P}_6\lambda\mathcal{P}_7.\mathcal{P}_6(\lambda z.\mathcal{P}_7(\mathit{arg}_1(z)));$ $\lambda\mathcal{P}_8.\mathcal{P}_8(\mathit{marko});$
|| $\text{NP}_{\text{GEN}} \Rightarrow \text{PP}_{\text{OD-PASS}}$ NP_{GEN}
|| --

$\varphi_3.[\text{zatvorila}\bullet\text{se}\bullet\text{knjiga}\bullet\varphi_3];$ ||
$\lambda\mathcal{P}_3.\exists(\mathbf{book})(\lambda y.\mathcal{P}_3(\lambda R.[\mathit{close}(\mathit{arg}_2(y) \cap R)]));$ ||
$\text{PP} \Rightarrow \text{S}$ od•marka; $\lambda\mathcal{P}_7.\mathcal{P}_7(\mathit{arg}_1(\mathit{marko}))$; $\text{PP}_{\text{OD-PASS}}$

||

zatvorila•se•knjiga•od•marka; $\exists(\mathbf{book})(\lambda y.[\mathit{close}(\mathit{arg}_2(y) \cap \mathit{arg}_1(\mathit{marko}))])$; S

As one can see, both the English and Croatian passive constructions have remarkably similar proofs and corresponding semantic interpretations. The next section will demonstrate the extension of this parallelism to two constructions which are seemingly different on the surface, but actually result from similar mechanisms, and arguably convey equally similar semantic interpretations; the Croatian impersonal ("*zatvorilo se knjigu* *[*od Marka*]"), and the English middle ("*a book closed* *[*by Bob*]") voices.

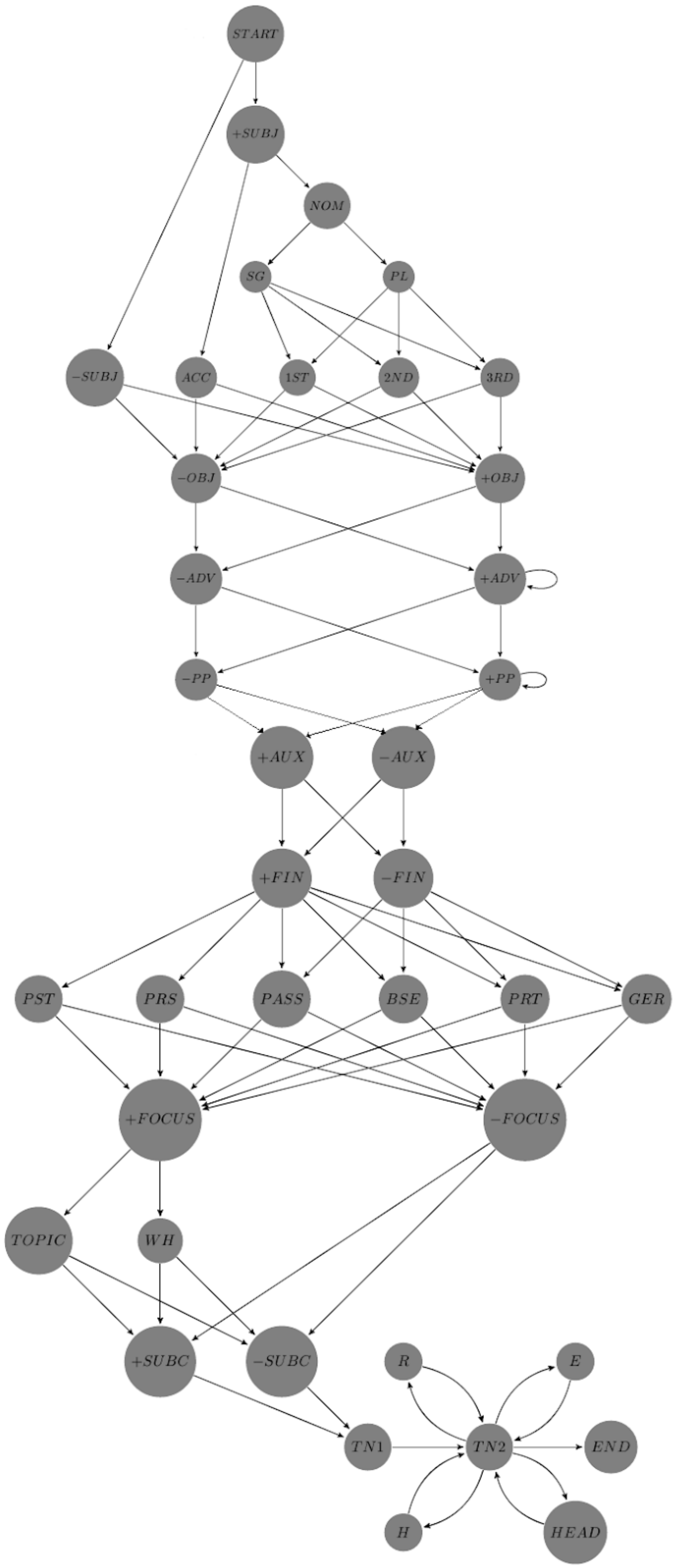
(2.3) Passive-Like Constructions

Both the Croatian impersonal and the English middle constructions fall into a category designated here as *passive-like*; they share many of the same characteristics with their respective "true" passive counterparts-- the *primary argument* is removed from its canonical position, there is a morphosyntactic variation from the canonical active construction, and a more oblique semantic argument is assigned to the syntactic subject argument, which is canonically assigned the *primary argument*. However, they both additionally display morphosyntactic variation from their respective "true" passive counterparts, and, unlike the passives, neither one allows the reintroduction of the *primary argument* via a prepositional phrase. The following two sections (2.3.1 and 2.3.2) will demonstrate that both passive-like constructions have similarities, but also large deviations, in their pre-syntactic derivational paths compared to those of their respective "true" passive counterparts. Semantically, they both display varying degrees of "distance" from their *primary arguments*; while the Croatian impersonal carries a negated existential interpretation (Rivero 2001) on its *primary*/agent-like semantic role (implying that no animate agent exists, i.e. that some external inanimate force caused the action), the English middle seems to display a weaker version of this phenomenon-- in two non-active sentences e.g. "*the boat sank*" vs. "*the boat was sunk*", the second (i.e. passive) seems to imply volition on behalf of some entity. In other words, "*the boat sank*" seems to prompt the question "*how?*", while "*the boat was sunk*" seems to prompt the question "*by whom?*".

(2.3.1) English Middle Voice

One of the two terms necessary derive the English sentence "*the book closed*", *the book*, has already been derived in section (2.1.1). As such, the only term left to derive is the middle form of the verb *close*-- morphologically identical to its active counterpart, the only overt distinguishing feature is its reduced valence; this form only allows a single argument (i.e. it is intransitive), as opposed to its transitive canonical counterpart. The *Map* for CLOSE, given in figures 2 and 11, is given once again in figure 15 below for reference:

Figure 15


START
+SUBJ
NOM
SG
PL
−SUBJ
ACC
1ST
2ND
3RD
−OBJ
+OBJ
−ADV
+ADV
−PP
+PP
+AUX
−AUX
+FIN
−FIN
PST
PRS
PASS
BSE
PRT
GER
+FOCUS
−FOCUS
TOPIC
WH
+SUBC
−SUBC
R
E
TN1
TN2
END
H
HEAD

As in the passive, this derivation proceeds in an identical fashion to that of the active form, until it reaches the +OBJ and -OBJ nodes. At this point, as in the passive, the system moves to -OBJ, and does not select an object. Unlike the passive, it then follows the path -ADV → -PP → -AUX → +FIN. As the boolean *auxiliary* = 0, PASS is not a valid move available to the automaton, and as such it proceeds to PST, as in the active. However, as it lacks an object, in the class of verbs that allow for a middle voice interpretation, to which CLOSE belongs, the subject may be assigned either the arg_1 or arg_2 roles, chosen nondeterministically by the system. More formally, the -OBJ node, which was not selected in the active derivation, does *not* remove an element from the *open_roles* set, which is then $\{arg_1, arg_2\}$ at the point that a role is assigned to the subject argument. In this case, arg_2 is chosen, and the rest of the derivation proceeds identically to the previous two in sections (2.1.1) and (2.2.1). The final term outputted by the automaton is then the following:

24. $\lambda\varphi.[\varphi\bullet\text{closed}]; \lambda\mathcal{P}.\mathcal{P}(\lambda x.[close(arg_2(x))]); NP_{NOM} \Rightarrow S$

Given this term and that already derived for *a book*, an extremely simple derivational proof is then possible:

Figure 16

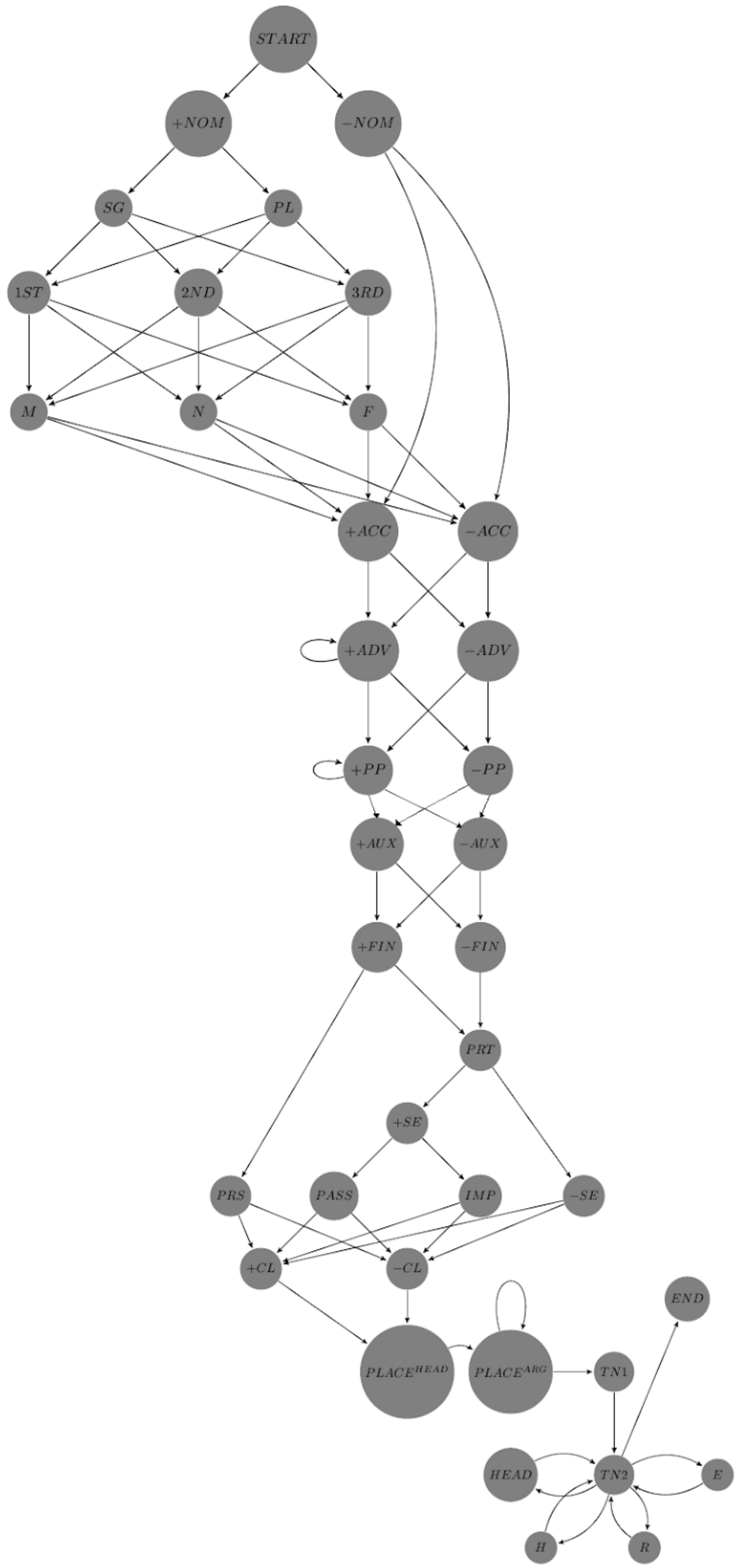
$$\dfrac{\lambda\varphi.[\varphi\bullet\text{closed}];\ \lambda\mathcal{P}_1.\mathcal{P}_1(\lambda x.[close(arg_2(x))]);\ NP_{NOM} \Rightarrow S \qquad \text{a}\bullet\text{book};\ \lambda\mathcal{P}_2.\exists(\mathbf{book})(\lambda y.\mathcal{P}_2(y));\ NP_{NOM,ACC}}{\text{a}\bullet\text{book}\bullet\text{closed};\ \exists(\mathbf{book})(\lambda y.[close(arg_2(y))]);\ S}\ ||$$

For a possible theoretical explanation as to why this construction (and other constructions lacking subjects e.g. nonfinite verb forms) cannot take a $PP_{BY\text{-}PASS}$ argument, see section (4.2). Regardless, this paper hypothesizes that it is precisely this inability to reintroduce a *primary argument* that results in its less-volitional interpretation compared to that of its active and passive counterparts-- following the *Gricean Maxim of Quantity* (Engelhardt et al. 2006), a speaker is more likely to utilize a construction in which it is impossible to introduce a *primary argument* if said argument is unknown or nonexistent.

(2.3.2) Croatian Impersonal

The Croatian impersonal constructions are essentially the stronger version of the English middle-- they confer the following interpretation; $\neg\exists(\mathbf{x})(\lambda y.P'(...arg_1(y)...))$, where P' is an arbitrary propositional predicate. As in the English middle, these constructions do not allow the reintroduction of the arg_1 argument via an *od*-phrase. Somewhat inexplicably, the main verb in these constructions does not agree with *any* NP argument within the phrase, instead taking on third person neuter singular agreement, regardless of the presence of an $NP_{CASE,3,N,SG}$, where CASE represents any arbitrary noun case. In order to derive the impersonal construction given in (3)-- "*zatvorilo se knjigu*"-- only the impersonal verb, *zatvorilo*, must be derived, as SE and *knjigu* are already given in earlier sections. The *Map* for ZATVORITI previously given in figures 6 and 13 is again given for reference in figure 17 below:

Figure 17


START
+NOM
−NOM
SG
PL
1ST
2ND
3RD
M
N
F
+ACC
−ACC
+ADV
−ADV
+PP
−PP
+AUX
−AUX
+FIN
−FIN
PRT
+SE
PRS
PASS
IMP
−SE
+CL
−CL
END
PLACE^HEAD
PLACE^ARG
TN1
HEAD
TN2
E
H
R

The course of this derivation immediately deviates from the previous active and passive derivations of ZATVORITI. From the START node, the system's first move is to -NOM, and as such it does not select any morphosyntactic subject agreement features. From -NOM, the system proceeds to +ACC, adding an accusative argument, before advancing along the following path; -ADV → -PP → +AUX → +FIN → PRT → +SE, paralleling its passive counterpart (with the exception of -PP). As in the passive derivation, +SE essentially overwrites the auxiliary verb from the *clitic* variable, replacing it with the SE clitic. From +SE, the system moves to IMP (impersonal), whose *condition function* is only satisfied if there exists an NP_{ACC} argument, and there does not exist an NP_{NOM} argument, the exact opposite of the PASS *condition function*. The *instruction function* of the IMP node sets $\varphi_{THIS} = \beta_{ZATVORITI}(P^{TCP} \bullet 3^{N} \bullet 5^{SG})$ (= "*zatvorilo*"), and $\Sigma_T = \neg\exists(\mathbf{x})(\lambda z.\mathit{close}(\text{ARG\_P} \cap \mathit{arg}_1(z)))$. In this case, ARG_P = $\mathit{arg}_2(y)$, where y is the semantic entity-type variable corresponding to the NP_{ACC} argument. From IMP, the rest of the derivation proceeds identically to the active and passive pre-syntactic derivations of ZATVORITI. At the end of stage 8, this results in the following proof calculus term:

25. $\lambda\varphi_1\lambda\varphi_2.[\text{zatvorilo}\bullet\varphi_1\bullet\varphi_2]; \lambda\mathcal{P}_1\lambda\mathcal{P}_2.\mathcal{P}_2(\lambda y.\mathcal{P}_1(\neg\exists(x)(\lambda z.[\mathit{close}(\mathit{arg}_1(z) \cap \mathit{arg}_2(y))])));$
$SE \Rightarrow NP_{ACC} \Rightarrow S$

Once again, this term, along with those corresponding to SE and *knjigu*, combine in an extremely straightforward derivational proof:

Figure 18

$\lambda\varphi_1\lambda\varphi_2.[\text{zatvorilo}\bullet\varphi_1\bullet\varphi_2];$
$\lambda\mathcal{P}_1\lambda\mathcal{P}_2.\mathcal{P}_2(\lambda y.\mathcal{P}_1(\neg\exists(\mathbf{x})(\lambda z.[\mathit{close}(\mathit{arg}_1(z) \cap \mathit{arg}_2(y))])));$
$SE \Rightarrow NP_{ACC} \Rightarrow S$ se; $\lambda\mathcal{P}_3.\mathcal{P}_3$; SE

| |

$\lambda\varphi_2.[\text{zatvorilo}\bullet\text{se}\bullet\varphi_2];$
$\lambda\mathcal{P}_2.\mathcal{P}_2(\lambda y.\neg\exists(\mathbf{x})(\lambda z.[\mathit{close}(\mathit{arg}_1(z) \cap \mathit{arg}_2(y))]);$ knjigu; $\lambda\mathcal{P}_4.\exists(\mathbf{book})(\lambda w.\mathcal{P}_4(w));$
$NP_{ACC} \Rightarrow S$ NP_{ACC}

| |

zatvorilo•se•knjigu; $\exists(\mathbf{book})(\lambda w.\neg\exists(\mathbf{x})(\lambda z.[\mathit{close}(\mathit{arg}_1(z) \cap \mathit{arg}_2(w))]))$; S

The non-agentive interpretation is much clearer in this construction compared to its English counterpart-- the agent-like argument is literally defined as nonexistent in the logical semantic interpretation. Again, for a possible theoretical explanation as to why both this and the

English middle constructions laid out in section (2.3.1) cannot reintroduce the *primary argument* in a prepositional phrase, see section (4.2).

(3.4) Deverbal Adjectival Forms

Both English and Croatian allow *deverbal adjectival* forms of CLOSE and ZATVORITI, respectively. These constructions convert the verb into an adjective, and in English, this verb form (at least for the verb *close*, there are exceptions) is identical to both the past tense and past participial forms-- "*closed*". On first glance, these constructions seem identical to the passive forms; compare the passive "<u>*the book was closed*</u> *by John*" and the deverbal adjectival "<u>*the book was closed*</u> *when John entered the room*". The key difference is in interpretation; in the present tense, the deverbal adjectival has a stative reading, while the passive carries a stative interpretation. Again, compare the passive "*the bar is closed by the health department every year, but it always manages to open up again*" and the deverbal adjectival "*nobody can enter, this bar is closed* $^{?}$[*by the health department*]"-- as in the middle construction, the deverbal adjectival form does *not* allow the reintroduction of arg_1 via a *by*-phrase, unlike the superficially identical passive.

In direct contrast, the Croatian deverbal adjectival form ("*knjiga je bila zatvorena* [*od Marka*]") *does* allow the reintroduction of arg_1 via an *od*-phrase. Additionally, the -*en* deverbal adjectival morpheme is entirely distinct from the participial -*il*, another difference from its English counterpart.

Unfortunately, the present analysis does not contain a satisfying account of the Croatian and English deverbal adjectival forms. To date, the most compelling account posits a node VTADJ $\in V_{CLOSE}, V_{ZATVORITI}$ ("verb-to-adjective"), such that for both the CLOSE and ZATVORITI lexemes, <START, VTADJ>, <VTADJ, END> $\in E_{CLOSE}, E_{ZATVORITI}$ (i.e. the only possible path that moves to VTADJ is START → VTADJ → END), that does not allow either form to take any arguments, and confers a semantic interpretation along the lines of λZ.*close*(Z). A second entry/*Map* path for the copular auxiliary forms of BE and BITI then applies arg_2 (and any other adverbial and PP-type arguments selected by said auxiliaries) to λZ.*close*(Z), and some mechanism allows BITI to combine with $PP_{BY\text{-}PASS}$, while preventing BE from doing the same.

Intransitives e.g. the English ARRIVE that allow synonymous constructions à la "*a man arrived*" and "*there arrived a man*" could potentially be analyzed as conferring the arg_2 role to their (non-expletive) subject arguments *canonically*. "Normal" adjectives e.g. green could also have the semantic form λx.*green*(arg_2(x)), allowing for a single auxiliary BE that can take as an argument any kind of deverbal adjective derived from a verb of any canonical valence, in addition to "regular/normal" (i.e. *lexical*) adjectives, as its object argument, and for any given NP subject with the semantic representation *x*, yield the following interpretation; $adj_i(\ldots arg_2(x)\ldots)$. Under this informal analysis, the following three constructions, which seem to assign different semantic roles, could all be accounted for with a singular lexical entry for the copular BE (26b, 27b, and 28b provide potential corresponding semantic interpretations):

26a. *The man is gone.*
b. $go(arg_2(\iota(man)))$

27a. *The car is ruined.*
b. $ruin(arg_2(\iota(car)))$

28a. *The house is white*.
b. $white(arg_2(\iota(house)))$

While this does seem to be an interesting avenue of theoretical inquiry, it also likely falls outside of the scope of this undergraduate thesis, and may be a better subject to pursue at a later date.

(4) Closing Remarks

Unfortunately, due to time constraints, there remain certain elements of this analysis and theoretical framework that were not able to be pursued as completely as I would have liked. The next two brief sections (4.1-4.2) lay out some potential routes of further research/inquiry, and attempt to clarify any lingering questions pertaining to the preceding sections of the paper.

(4.1) Lack of Isomorphism Between Semantic and Tectogrammatical Types

As a more formalism-minded reader may have noticed, many non-functional (i.e. non-implicational) tectogrammatical types have corresponding semantic types of the form $\lambda\mathcal{P}.[P]$. This clearly violates an established mantra of type-logical grammars-- a functional X-type must also have functional Σ- and φ-types, and vice-versa, a term with an irreducible X-type must have irreducible Σ- and φ-types as well. Unfortunately, this violation is necessary in order to allow certain uniformities such as the identical (tectogrammatical) treatment of quantified- and non-quantified-NPs. In fact, in this system, all types that do not have an $S_{FIN,\neg SUBC}$ (i.e. finite, non-subordinate clauses) *terminal reducing type* will inevitably have a Σ-sector of the form $\lambda\mathcal{P}.[P]$, even once they have reduced down to their respective *terminal reducing types*.

While it is admittedly unconventional, the benefits arguably outweigh the disadvantages in terms of the ability to (tectogrammatically) uniformly treat two or more terms with (seemingly) identical X-types, but differing Σ-types (e.g. quantified- and non-quantified-NPs). In a sense, all terms that do not have an X-type $S_{FIN,\neg SUBC}$ are “arguments waiting for a functor”, and in fact, partially as a consequence of this asymmetry, the set of grammatical constructions is exactly the set of terms whose Σ-types are well-formed truth-conditional logical formulae.

(4.2) LIMIT Sets

In section (2.2), the question remains unanswered; *what prevents $PP_{BY\text{-}PASS}$ from being an argument of non-passive constructions?* LIMIT sets may provide a solution. A LIMIT set (*global variable*) would be added at the START node of a given lexeme L_i, and is the set of all types forbidden to be arguments of L_i. If a given LIMIT set = e.g. $\{PP_A, PP_B, PP_C\}$, each PP argument added would then be of the form $PP_{\neg A, \neg B, \neg C}$, disallowing PPs of subtypes included in the LIMIT set. $PP_{BY\text{-}PASS}$ would then be an element of essentially every LIMIT set, and the i*nstruction function* of the PASS node would then contain a substring of the following form;
LIMIT = LIMIT \ $\{PP_{BY\text{-}PASS}\}$.

(4.3) Clitic Clusters

While each of the respective examples in this paper only contains a single clitic, SE or *je*, Croatian permits clusters of multiple clitics e.g. *mi je* (*mi* = first person singular dative pronoun), which are required to appear in clitic position in a sequence uninterrupted by any intervening material (Schütze 1994). In order to account for this phenomenon, the variable *clitic*, which appears in the present analysis as a single $<\varphi, \Sigma, X>$ triple, could instead be realized as an n-tuple; a well-ordered sequence of $<\varphi, \Sigma, X>$ triples, each representing a different clitic in the chain. The +SE node would then be instructed to only overwrite specific ordinal positions or *tectogrammatical* types within this sequence (or some combination of the two). A comprehensive NECG analysis of Croatian clitic clusters, however, is outside of the scope of this thesis, although an account of this data represents a possible prospective avenue for further research.

(5) Conclusion

In section (1.1), three main variations between Croatian active, passive, impersonal, and deverbal adjectival constructions, specifically applied to the transitive verb ZATVORITI, are identified. First, the absence of the auxiliary clitic *je* in the impersonal and passive was analysed as the +SE node (which must be visited in order to access the IMP and PASS nodes) overwriting the *clitic* variable, which was at that point occupied by the auxiliary *je*, with a term of the form $<\varphi_i, \mathcal{P}_j, SE>$.

This *clitic*-deletion process accounts for the phenomenon which motivated earlier analyses of an underlying, phonologically unrealized auxiliary clitic (Zec 1985); the absence of an overt finite verb in the passive and impersonal constructions. The sequence of steps necessary for the automaton to *Pre-Syntactic Automaton* to select for the auxiliary clitic will set its *terminal reducing type* to S_{FIN}, while also deriving a participial (nonfinite) morphophonological form of ZATVORITI. The +SE node, which allows for the impersonal and passive forms, overwrites the *je* clitic, but makes no reference or changes to the derivation's current *terminal reducing type* or morphosyntactic feature specification, effectively allowing a term with nonfinite morphology to yield a valid finite construction.

Secondly, the lack of morphosyntactic agreement inherent in the impersonal construction is explained by the fact that from the START node, the first move necessary to derive an impersonal construction is to -NOM -- bypassing the agreement nodes SG, PL, 1, 2, 3, M, N, and F. Finally, the impersonal construction's non-agentive interpretation is analysed purely as result of the intersection between this lack of morphosyntactic specification, and its selection of the *non-canonical marker*, SE. The n^{th}-Order polymorphic Σ-variable, $\mathcal{P}$, allows SE to act as an identity function ($\lambda\mathcal{P}.\mathcal{P}$) for whichever construction may select for it, giving the language the flexibility to use SE in a wide variety of roles. Croatian verbs of the same class as ZATVORITI are hypothesized to select for SE to mark a construction's inability to assign its agent-like role to a syntactic argument (note that the possibility of introducing the agent-like argument via *od*-phrase in the passive is a feature of the lexeme OD, rather than the passive construction itself). The lack of a nominative NP argument in the impersonal is then an epiphenomenon stemming from its lack of morphosyntactic subject agreement features.

In terms of the formalism, *Nondeterministically Enumerated Categorial Grammar* allows for precise yet powerful control over every aspect of a given phrasal structure, before the actual syntactic derivation even begins. It additionally provides an extremely simple, and, perhaps more importantly, predictable proof calculus that permits only one Rule of Inference: (nondirectional) Implication Elimination. However, this simplicity does not prevent the formalism from accounting for scopal ambiguities, nor, as section (4.3) briefly touches upon, does it necessarily rule out the possibility of (albeit indirectly) referencing overt linear order and directionality, by virtue of the *left-to-right normal form* imposed via the *Pre-Syntactic Automaton* on all proof terms.

In conclusion, the NECG framework provides an exciting new formal perspective for theoretical syntactic analysis, and while this paper is in no way proof of its cross-linguistic viability and utility, it (hopefully) demonstrates the flexibility and potential analytic utility of the system.

(6) Naming Conventions For Typed Variables

The following section provides a quick-reference guide that enumerates the naming conventions used in this paper for variables over each set-theoretic, semantic, tectogrammatical, and phonological type.

(6.1) Set-Theoretic (& Miscellaneous) Types

-Z: *Unordered Sets*

-W: *Well-ordered Sets/Tuples*

-{i, j, k, m, n}: *Positive and Negative Integers*

(6.2) Semantic Types

-{u, v, w, x, y, z}: *Entities*

-R: *Functional Predicates* (i.e. functions of type R: Z → Z)

-P: *Relational Predicates* (i.e. functions of type R: Z → {1, 0})

-$\mathcal{P}$: *Polymorphic over all n^{th}-Order Types and λ-Terms*

(6.3) Tectogrammatical Types

-X^S: *S-reducible Types*

-X^C: *Complex Types* (i.e. of the form $X_1 \Rightarrow X_{i>1}$)

-X^T: *Terminal Types* (any tectogrammatical type *not* of type X^C)

-X: *Polymorphic Over All X^T-, X^C-, and X^S-Types*

(6.4) Phonological Types

-φ: *Strings*

-σ: *String-to-String Function*

-$\boldsymbol{e}$: *Polymorphic Over φ- and σ-Types*

-β: *Paradigm Functions* (i.e. β: $\mathbf{N} \rightarrow \boldsymbol{e}$)

(7) Bibliography

1. Austin, Peter. "Causative and applicative constructions in Australian aboriginal languages." *The dative and related phenomena* (2005)

2. Baker, Mark, Kyle Johnson, and Ian Roberts. "Passive Arguments Raised", Linguistic Inquiry (1989).

3. Bozsahin, Cem. "Gapping and word order in Turkish." Proceedings of the 10th International Conference on Turkish Linguistics. 2000.

4. Comrie, Bernard, Martin Haspelmath, and Balthasar Bickel. "The Leipzig Glossing Rules: Conventions for interlinear morpheme-by-morpheme glosses." *Department of Linguistics of the Max Planck Institute for Evolutionary Anthropology & the Department of Linguistics of the University of Leipzig. Retrieved January* 28 (2008): 2010.

5. Curry, Haskell B. "Some logical aspects of grammatical structure." *Structure of language and its mathematical aspects*12 (1961)

6. De Groote, Philippe. "Towards abstract categorial grammars." *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*. 2001.

7. Dixon, R.M.W. & Alexandra Y. Aikhenvald (eds) 2010. Case studies in transitivity. Cambridge: Cambridge University Press.

8. Engelhardt, Paul E., Karl GD Bailey, and Fernanda Ferreira. "Do speakers and listeners observe the Gricean Maxim of Quantity?." Journal of Memory and Language 54.4 (2006): 554-573.

9. Frick, Markus, and Martin Grohe. "The complexity of first-order and monadic second-order logic revisited." *Annals of pure and applied logic* 130.1-3 (2004)

10. Jansen, Jan Martin. "Programming in the λ-calculus: From Church to Scott and back." The Beauty of Functional Code. Springer, Berlin, Heidelberg, 2013. 168-180.

11. Kordić, Snježana. *Serbo-Croatian*. Lincom Europa, 1997.

12. Lambek, Joachim. "The mathematics of sentence structure." *The American Mathematical Monthly* 65.3 (1958): 154-170.

13. Moortgat, Michael. “Typelogical Grammar.” Stanford Encyclopedia of Philosophy, Stanford University, 7 Sept. 2010, plato.stanford.edu/entries/typelogical-grammar/.

14. Oehrle, Richard T. "Term-labeled categorial type systems." *Linguistics and Philosophy* 17.6 (1994): 633-678.

15. Progovac, Ljiljana. "Determiner phrase in a language without determiners (with apologies to Jim Huang 1982)." Journal of linguistics 34.1 (1998): 165-179.

16. Rice, Keren. "Principles of affix ordering: An overview." Word Structure 4.2 (2011): 169-200.

17. Rivero, María Luisa, and M. Milojević Sheppard. "On Impersonal se/się in Slavic." Current Issues in Formal Slavic Linguistics. Peter Lang, Frankfurt/Main (2001): 137-147.

18. Schütze, Carson. "Serbo-Croatian second position clitic placement and the phonology-syntax interface." MIT working papers in linguistics 21 (1994): 373-473.

19. Spalatin, Leonardo. 1973. Some Serbo-Croatian equivalents of the English passive. The Yugoslav Serbo-Croatian-English Contrastive Project Reports 8.115-31.

20. Stewart, Thomas, and Gregory Stump. "Paradigm function morphology and the morphology-syntax interface." *Linguistic interfaces* (2007)

21. Sørensen, Morten Heine, and Pawel Urzyczyn. *Lectures on the Curry-Howard isomorphism*. Vol. 149. Elsevier, 2006.

22. Zec, Draga. "Objects in Serbo-Croatian." Annual Meeting of the Berkeley Linguistics Society. Vol. 11. 1985.